

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AGARD-CP-360

AGARD-CP-360

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT

11, rue de l'Aviation - 92230 Neuilly-sur-Seine - France

AD-A154 327

AGARD CONFERENCE PROCEEDINGS No.360

Cost Effective and Affordable Guidance and Control Systems

DTIC
ELECTE
S JUN 3 1985 D
A

This document has been approved
for public release and sale; its
distribution is unlimited.

DTIC FILE COPY

NORTH ATLANTIC TREATY ORGANIZATION



DISTRIBUTION AND AVAILABILITY
ON BACK COVER

85 5 28 069

AGARD-CP-360

NORTH ATLANTIC TREATY ORGANIZATION
ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT
(ORGANISATION DU TRAITE DE L'ATLANTIQUE NORD)

AGARD Conference Proceedings No.360
**COST EFFECTIVE AND AFFORDABLE
GUIDANCE AND CONTROL SYSTEMS**

Papers presented at the Guidance and Control Panel 39th Symposium held
at the Yunus Holiday Resort Hotel, Cesme, Izmir, Turkey, 16-19 October 1984.

THE MISSION OF AGARD

The mission of AGARD is to bring together the leading personalities of the NATO nations in the fields of science and technology relating to aerospace for the following purposes:

- Exchanging of scientific and technical information;
- Continuously stimulating advances in the aerospace sciences relevant to strengthening the common defence posture;
- Improving the co-operation among member nations in aerospace research and development;
- Providing scientific and technical advice and assistance to the North Atlantic Military Committee in the field of aerospace research and development;
- Rendering scientific and technical assistance, as requested, to other NATO bodies and to member nations in connection with research and development problems in the aerospace field;
- Providing assistance to member nations for the purpose of increasing their scientific and technical potential;
- Recommending effective ways for the member nations to use their research and development capabilities for the common benefit of the NATO community.

The highest authority within AGARD is the National Delegates Board consisting of officially appointed senior representatives from each member nation. The mission of AGARD is carried out through the Panels which are composed of experts appointed by the National Delegates, the Consultant and Exchange Programme and the Aerospace Applications Studies Programme. The results of AGARD work are reported to the member nations and the NATO Authorities through the AGARD series of publications of which this is one.

Participation in AGARD activities is by invitation only and is normally limited to citizens of the NATO nations.

The content of this publication has been reproduced directly from material supplied by AGARD or the authors.

Published February 1985

Copyright © AGARD 1985
All Rights Reserved

ISBN 92-835-0373-2



Printed by Specialised Printing Services Limited
40 Chigwell Lane, Loughton, Essex IG10 3TZ

11

PREFACE

Advanced manned and unmanned air-vehicles have to comply with sophisticated performance requirements. For this reason, the corresponding guidance and control systems are becoming increasingly complex. In addition, modern manned and unmanned weapon systems are assuming an increasing tactical/operational importance and demand higher percentages of the relevant budgets. Procurement agencies are thus faced with the problem of rapidly increasing costs for urgently needed systems. If this trend continues, new systems will not be affordable in required quantities.

On the other hand, considerable progress has been made in relevant technologies: signal-processing (hardware and software), low-cost sensors, means for positioning and updating, etc. Further advances can be expected in the future.

For this purpose, the Guidance and Control Panel is directing its efforts towards the application of available technology to the design, operation, maintenance and testing of cost-effective and affordable guidance and control systems.

In order to perform this objective, the symposium has been arranged as follows: cost-effectiveness models; system configurations and design tools; low-cost guidance and control sensors; computational techniques and data processing; reduction of development and support costs; examples of cost-effective achievements and approaches.

The symposium concluded with a Round Table Discussion in which Panel members, speakers, and other attendees participated.

Les véhicules aériens, pilotés ou non, doivent se conformer à des exigences complexes de performance. De ce fait, les systèmes correspondants de guidage et de pilotage deviennent de plus en plus sophistiqués. En outre, les systèmes d'armes, pilotés ou non, revêtent une importance grandissante aux plans tactique et opérationnel, et requièrent, des budgets appropriés, des pourcentages plus élevés. Les organismes chargés de leur fourniture sont donc confrontés au problème d'un accroissement des coûts pour des systèmes dont la nécessité s'impose avec urgence. Si cette tendance se poursuit, il ne sera pas possible de se procurer les nouveaux systèmes en nombre suffisant.

D'autre part, des progrès considérables ont été réalisés dans les domaines technologiques concernés: traitement des signaux (matériel et logiciel), détecteur de coût modique, moyens de positionnement et de mise à jour, etc. On peut escompter que d'autres progrès seront réalisés à l'avenir.

Dans ce but, le Panel de Guidage et Pilotage centre ses efforts sur l'application de la technologie existante à la conception, au fonctionnement, à la maintenance et aux essais de systèmes de guidage et de pilotage présentant un rapport coût-efficacité satisfaisant et accessible aux usagers.

Pour répondre à cet objectif, le symposium a été divisé suivant différents thèmes: modèles de coût-efficacité; configurations des systèmes et instruments conceptuels; détecteurs de coût modique pour le guidage et le pilotage; techniques de calcul et de traitement de données; réduction des coûts de soutien et de développement; exemples de réalisations et d'approches présentant un rapport coût-efficacité satisfaisant.

Le symposium s'est terminé par une "Table Ronde" à laquelle ont pris part les membres du Panel, les conférenciers et les autres participants.



| | |
|--------------------|-------------------------------------|
| Accession For | |
| NTIS GRA&I | <input checked="" type="checkbox"/> |
| DTIC TAB | <input type="checkbox"/> |
| Unannounced | <input type="checkbox"/> |
| Justification | |
| By _____ | |
| Distribution/ | |
| Availability Codes | |
| Dist | Avail and/or Special |
| A-1 | |

GUIDANCE AND CONTROL PANEL OFFICERS

Chairman: Dr Ing. R.C.Onken
DFVLR
Institut für Flugführung
Postfach 32 67
D-3300 Braunschweig
GE

Deputy Chairman: Mr K.A.Peebles
Director, Technology Application
Communication Electronics
National Defence Headquarters
Ottawa, Ontario, K1A 0K2
Canada

TECHNICAL PROGRAMME COMMITTEE

Chairman: Mr U.Krogmann, GE
Members: Dr A.Benoit, BE
Mr G.Bonnevie, FR
Ing. P.L.Ferraris, IT
Mr A.R.Kazakoğlu, TU
Mr J.L.Hollington, UK
Dr W.P.Albritton, US

PANEL EXECUTIVE

Mr B.M.Heliot
AGARD-NATO
7 rue Ancelle
92200 Neuilly-sur-Seine
France

HOST NATION COORDINATOR

Mr A.R.Kazakoğlu
TÜBİTAK-BAE
Atatürk Bulvarı 221
Kavaklıdere
Ankara
Turkey

ACKNOWLEDGEMENT/REMERCIEMENT

The panel wishes to express its thanks to the Turkish National Delegates to AGARD for their invitation to hold this meeting in Çesme, Izmir, Turkey, and for the facilities and personnel which made the meeting possible.

Le Panel tient à remercier les Délégués Nationaux de la Turquie près l'AGARD de leur invitation à tenir cette réunion à Çesme, Izmir, Turquie, ainsi que pour les installations et le personnel mis à sa disposition.

CONTENTS

| | Page |
|---|-----------|
| PREFACE | iii |
| PANEL OFFICERS AND PROGRAMME COMMITTEE | iv |
| TECHNICAL EVALUATION REPORT by C. Baron | vii |
| | Reference |
| REVIEW OF MISSION VERSUS COST REQUIREMENTS AND TECHNICAL ISSUES — Keynote Address — by C. Baron | K |
| <u>SESSION I — COST EFFECTIVENESS MODELS/SYSTEMS CONFIGURATION AND DESIGN TOOLS</u> Chairman: Dr A. Benoit, BE | |
| SOME ASPECTS OF HOW TO DESIGN COST-EFFECTIVE FLIGHT CONTROL SYSTEMS by U. Butter and L. Botzler | 1 |
| ANALYSE COMBINATOIRE PERFORMANCES/COUTS D'UN SYSTEME AUTONOME DE NAVIGATION POUR AERONEFS par P. Lloret et G. Lavoipierre | 2 |
| NAVIGATION — ACCOUNTING FOR COST by D. J. Hamlin | 3 |
| DESIGN ADEQUACY: AN EFFECTIVENESS FACTOR by A. R. Habayeb | 4 |
| THE EVOLUTIVE STATE: A NEW DEFINITION OF SYSTEM STATE TO INCLUDE SEQUENCING OF EVENTS by V. Amoia and R. Somma | 5 |
| OPTIMISATION DES SPECIFICATIONS D'UN SYSTEME AIR/SURFACE LASER par J. P. Murgue et D. Evrard | 6 |
| READINESS THROUGH ENVIRONMENTAL STRESS SCREENING AT LOW OR NO COST INCREASE* by J. L. Capitano | 7 |
| DESIGN-TO-COST (DTC) METHODOLOGY TO ACHIEVE AFFORDABLE AVIONICS by A. J. Shapiro | 8 |
| <u>SESSION II — LOW COST, HIGH RELIABILITY GUIDANCE AND CONTROL SENSORS</u> Chairman: Mr G. Bonnevie, FR | |
| PREGUIDAGE ET PILOTAGE A L'AIDE DE SENSEURS INERTIELS COMPOSITES par J. Resseguier | 9 |
| LASER GYROSCOPE RANDOM WALK DETERMINATION USING A FAST FILTERING TECHNIQUE by J. G. Mark and A. Brown | 10 |
| THE SERIES 2000 — A MINIATURE GAS BEARING DTG FOR LOW COST STRAPDOWN GUIDANCE AND CONTROL by G. Beardmore | 11 |
| SYSTEME DE GUIDAGE A GYROLASER POUR MISSILES A MOYENNE PORTEE par B. de Salaberry | 12 |

*Not available at time of printing

| | Reference |
|---|-----------|
| UTILISATION D'UN MAGNETOMETRE AUTOCOMPENSE DANS UN SYSTEME DE NAVIGATION ECONOMIQUE POUR HELICOPTERE par J.L.Roch, J.C.Goudon et Ph.Chaix | 13 |
| <u>SESSION III – COMPUTATIONAL TECHNIQUES AND DATA PROCESSING</u> | |
| Chairman: Dr H.Sorg, GE | |
| A COST-EFFICIENT CONTROL PROCEDURE FOR THE BENEFIT OF ALL AIRSPACE USERS by A.Benoit and S.Swierstra | 14 |
| WIND MODELLING FOR INCREASED AIRCRAFT OPERATIONAL EFFICIENCY by W.Lechner and R.Onken | 15 |
| A METHOD OF ESTIMATING AIRCRAFT ATTITUDE FROM FLY BY WIRE FLIGHT CONTROL SYSTEM DATA BY R.J.V.Snell | 16 |
| THE IMPACT OF VLSI ON GUIDANCE AND CONTROL SYSTEM DESIGN by D.Price | 17 |
| REUSABLE SOFTWARE – A CONCEPT FOR COST REDUCTION by C.Anderson and M.Henne | 18 |
| MAINTAINING CONSISTENCY AND ACCOMMODATING NETWORK REPAIR IN A SELF-REGENERATIVE DISTRIBUTED DATABASE MANAGEMENT SYSTEM by M.Reilly, P.R.Tillman, R.Catt and R.B.Moore | 19 |
| <u>SESSION IV – REDUCTION OF DEVELOPMENT AND SUPPORT COSTS</u> | |
| Chairman: Mr J.L.Hollington, UK | |
| SIMULATION – A TOOL FOR COST-EFFECTIVE SYSTEMS DESIGN AND LIFE TEST REDUCTION by R.Gauggel | 20 |
| DESCRIPTION DU PROCESSUS D'ELABORATION DU DIALOGUE "HOMME-MACHINE" DANS LES AVIONS D'ARMES EQUIPES DE TUBES CATHODIQUES; ROLES ET DESCRIPTIVE DE MOYENS DE SIMULATION PILOTEE par P.Helie | 21 |
| INTEROPERABILITY THROUGH STANDARD AIRCRAFT/WEAPON INTERFACES* by P.Cunningham and A.M.Henne | 22 |
| <u>EXAMPLES OF COST-EFFECTIVE ACCOMPLISHMENTS AND APPROACHES</u> | |
| Chairman: L.Ferraris, IT | |
| TERRAIN FOLLOWING WITHOUT USE OF FORWARD LOOKING SENSORS by H.F.Schwegler and B.Schreiner | 23 |
| THE USE OF PRESSURE SENSING TAPS ON THE AIRCRAFT WING AS SENSOR FOR FLIGHT CONTROL SYSTEMS by D.Brunner | 24 |
| LOW COST TWO GIMBAL INERTIAL PLATFORM AND ITS SYSTEM INTEGRATION by R.N.Priestley and J.B.Towler | 25 |
| COST-EFFECTIVE MISSILE GUIDANCE THROUGH DISTRIBUTED ARCHITECTURE AND STANDARD PROCESSORS by H.Gloerud | 26 |

*Not available at time of printing

TECHNICAL EVALUATION REPORT

by

C. Baron, M.Sc.
Head, CM (s) 1/SAS
Room 221 Ministry of Defence
Savoy Hill House, Savoy Hill
London, UK

1. INTRODUCTION

The 39th Symposium of the Guidance and Control Panel of AGARD was held at the Altin Yunus Hotel, Çeşme, Turkey, from the 16th to 19th October, 1984. The Programme Chairman was Mr U.K. Krogmann of Bodenseewerk, FRN-EN, Überlingen, Federal Republic of Germany. The meeting was attended by upwards of 110 people, among whom the US, UK, FRG, France and the host nation Turkey had relatively large and roughly comparable participation: 12 nations in all being represented.

2. THEME AND OBJECTIVES

The steady and universal increase in defence equipment costs documented in the Keynote Address was relatively easy for national economies to bear in the period of high economic growth following World War II, but more recently, lower growth has brought increasing difficulty for most nations even to maintain forces at current strengths. Conscious as they were of the urgent need to tackle this most important problem, it was nevertheless a bold step by the Guidance and Control Panel to organise this symposium, for the drive to reduce costs is mistakenly regarded by too many in the scientific and technical community as an unwelcome restraint on progress. The part played by this meeting in dispelling this impression must have been sufficient justification in itself. It is easy to show that if participation in the conference generated a single idea capable of saving 1% of the cost of a major defence programme, that saving could exceed the total cost of the symposium by two orders of magnitude.

The theme of the meeting was stated by the organisers as "the application of available technology to the design, operation, maintenance and testing of cost-effective and affordable guidance and control systems". It was particularly appropriate that this theme was thus developed in a country with a strong interest in low cost equipment. In addition a Special Session was organised, at the request of Turkey, on 'Strapdown gyros and accelerometers', which was highly appreciated by the Turkish participants; since it was not part of the symposium, however it will not be further referred to here.

3. TECHNICAL CONTENT

The meeting opened with an Address of Welcome by Col Kaya of the Turkish Air Force, who emphasised his nation's interest in, and welcome for the meeting. This was followed by the Keynote Address, given by the writer, which sought to establish beyond doubt the vital need for adopting all possible means of reducing costs, including harnessing technological innovation for the purpose.

Session I Cost effectiveness models/systems configuration and design tools.

In the first paper (1) Butter showed that, since maintenance accounts for by far the greatest proportion of life cycle costs, much thought, and if necessary extra cost, should be given in the development phase to reducing maintenance costs. Reviewing the various aspects of flight control system design, he indicated several approaches to achieve this. This was a well-presented and thoughtful paper which gave the conference a good start and provoked some active discussion which emphasised the need to avoid unnecessary over-specification.

Given the very wide range of costs of inertial navigation systems, depending on the accuracy required, it may be possible to design a cheaper system overall by augmenting IN information with data from other autonomous sources, such as magnetic or Doppler. Lavoipierre (2) introduced a formalised technique for arriving at the optimum cost solution, using charts relating performance parameters to cost, and allowing the differing characteristics of the different sensors to be accounted for.

Hamlin (3) also dealt with navigation systems, though on a broader basis, giving an interesting review of the options and their cost implications, while warning that breakthroughs in this well developed field seem unlikely; the most notable recent development had been the considerable improvement in inertial sensor reliability arising from the introduction of laser gyros.

In the unfortunate absence of the next scheduled speaker, Mr M.A. Ostgaard, Chief Scientist, AFWAL/FIG, Wright Patterson AFB, Dayton, Ohio, USA, nobly volunteered to speak on a related topic. His many friends were less than surprised to find that he happened to have a suitable set of vugraphs with him. Since his presentation was impromptu it is not reproduced in the proceedings and a more detailed resumé is therefore given here. He gave what was essentially a report of progress in several Flight Dynamics Lab programmes in the system integration and flight control field. At present the pilot has the task of integrating the many sub systems in the aircraft, notably flight control, propulsion, fire control and navigation, leading to his being overloaded, particularly in crucial combat or failure situations, with adverse effects both on capability and safety. The experimental Multi Function Flight Control Reference System (MFCRS) installed in an F15A aircraft is one step being taken towards greater integration, aiming to replace the separate inertial sensors of the navigation, flight control, and fire control systems by an integrated set meeting all the separate requirements. The system is based on two sets of three skew-mounted ring laser gyros, whose reliability is largely responsible for a 21 % projected reduction in life cycle costs. Other advantages are reduced weight (by 25%) and volume, 80% improvement in survivability and 50% in reaction time. The system is now installed and in flight test. The two gyropacks are separated for survivability, one being placed close to the radar to minimise foresighting errors, and the other behind the pilot's seat. This displacement, under bending of the fuselage, had caused discrepancies making it necessary to raise the trip levels registering disagreement of the sensors, so as to reduce false alarms in redundancy management; however, it had been demonstrated that faults would still be correctly identified and isolated. Flight tests had shown acceptable flight control at low dynamic pressures, but low damping was evident in all three axes on going to higher dynamic pressures; this will require the introduction of a better structural model. So far the conclusion is that the concept is very promising but the system needs more detailed development; reliability in practice had been impressive. The second programme described as the application of the ADA high level language to the F15 digital flight control system, which has reached 35 k ft and M 0.9, to 50 k ft and M 2.2. The system is a simple replacement of the original analogue computers by digital computers, in fact Zilog Z 8002's. In comparison with floating point assembly language it had been found that either Pascal or ADA required a memory expansion by 1.30, and time expansion by 1.10. However, programming hours required were in the ratio fixed point assembly language 1.0: floating point assembly language 0.4: ADA 0.2. Thus in both of the programmes described, key stages on the road to greater system integration were shown to offer real cost benefits.

The next two papers (Habayeb, 4 and Amoia and Somma, 5) both described somewhat academic approaches to the estimation/calculation of system effectiveness. The latter paper, delivered by Somma raised the interesting question of reliability estimation where the order in which internal failures occur may affect the outcome, and showed how to account for this.

In a return to the practical, Murgue and Evrard (6) jointly described the design of an air-launched laser guided glidebomb system for the attack of missile launching fast patrol boats. A particularly interesting feature was the extent to which designers had succeeded in interacting with the French Air Staff to ensure that the requirements were adjusted to allow the most cost-effective solution; a situation envied by many equipment designers in the audience. The presentation was illustrated by films showing the aircraft displays in operation, and the successful flight trials by the bomb.

The next presentation (Capitano, 7) was an object lesson in how not to address a multi lingual audience, the rate of delivery making interpretation impossible, and there were far too many slides. To those understanding English, however, it was an impressive and indeed scintillating talk. Asserting that the large proportion of defence budgets currently devoted to spares provision could be reduced by a factor of four by manufacturing techniques to improve reliability, he went on to show how to do it. The key elements were high thermal and vibrational stressing at the lowest possible level of assembly, intensive use of diagnostic equipment (electron microscopes etc) to detect component failures and find their causes, and effective feedback. Asked the cost of the diagnostic equipment required he said it was considerable but well justified by production cost savings; however good staff to operate it — "electronic pathologists" were very scarce.

In the last paper of the Session, Shapiro (8) described the formal implementation of Design-to-Cost (DTC) methodology to the development of Doppler navigation equipment. In addition to the necessary organisation procedures, Computer Aided Engineering (CAE) was shown to be an important aid. An active discussion period clearly demonstrated the interest aroused by this paper. Attention focussed on the costs and benefits of implementing such a system; it was stated that it had taken a year of interchange with the customer to establish the cost and performance requirements, after which the procedure had probably added 30% to the development time, but against this it was estimated that there had resulted a 30% reduction in unit production cost. The CAE system used had been a fairly simple one which had taken about 4 months to get into action, but the author would now advocate a more advanced system which he estimated might cost 1—1.5 M \$ and which would take 6—12 months to implement; one of his colleagues however thought the cost might be doubled, but both agreed that this would be well worth while provided it was used with the DTC methodology described. Stress was laid on the benefit arising from the stimulation of continuous interaction between customer and developer to ensure that requirements were agreed which would lead to the lowest cost equipment capable of satisfying the operational need.

Session II — Low cost, high reliability guidance and control sensors

This session brought together an interesting set of papers illustrating a variety of approaches to the design of low cost

sensors. First, Resseguier (9) described a miniature inertial navigation system for missile mid-course guidance, with low cost and volume and quick reaction as the prime features. Aiming to reduce inertial components to a minimum and compensate for resulting imperfections by computing, the system was based on the use of three ingenious 2-axis dry tuned rate gyros from which, by displacement of the centre of mass from the axis of rotation, linear acceleration could also be obtained. Mark (10) then described a technique for greatly reducing the cost of testing ring laser gyros, in which quantisation noise normally dominates the random walk errors, so that measurement of the latter requires very long observation periods. The fast filter method described can reduce this time from 3 hours to 2 minutes and is considered to give more reliable results.

Conflicting requirements have hitherto prevented the application of the valuable characteristics of the gas bearing to dry tuned gyros. Beardmore (11) showed how the difficulties could be overcome by clever design to produce a gyro with good intermediate range performance at low cost. The design is believed capable of extension to IN system performance or to even smaller size. In answer to questions it was stated that speed adjustment is used for tuning, only a narrow range of adjustment being necessary with electronic compensation. In the next paper (12), given by Salaberry, an alternative approach to low cost mid course missile guidance was described, in using three ring laser gyros and three accelerometers in an orthogonal strapdown configuration with mechanical dither. The paper demonstrates the widening application of laser gyros, and speculates on future advances. Another ingenious approach to navigation, this time for helicopters, was presented by Chaix (13); the outputs of a 3-axis magnetometer are resolved by a vertical gyro and combined with doppler ground speed, making extensive use of a computer for resolution compensation and calibration. This demonstration of how a set of well recognised elements can be assembled to give a new, elegant and low cost solution aroused considerable interest, discussion centring on the ease of flight test calibration of the system, the extent to which its use may be limited to areas of known magnetic variation, and possible future improvements.

This was a session of considerable interest which showed that designers in the navigation field are well aware of the need to reduce costs, and are responding with expertise and ingenuity.

Session III – Computational Techniques and Data Processing

The first paper of this session (Benoit 14) described an investigation of modified air traffic control procedures to provide greater aircraft fuel economy. Simulation and in-flight experiments showed considerable promise of accuracy of control of time of arrival, hence less fuel wastage in the terminal area, combined with economy in the cruise and descent phases. The paper gave rise to a number of questions.

The previous presentation having shown that lack of knowledge of the wind profile over the aircraft trajectory remains a major limitation to achieving better economy, it was appropriate that the next paper (Onken 15) should tackle this question, and it was tackled convincingly. The technique described uses known theoretical wind models in which the parameters are refined by Kalman filtering, using data measured in the aircraft and at the landing site; very good results had been achieved in flight tests, with satisfactory prediction ahead of the aircraft.

The redundant sensors of a fly-by-wire flight control system are a potential source of highly reliable information, and Snell (16) described how this could be used to provide a measurement of aircraft attitude. Originally conceived as a replacement for standby instruments and their displays, it was shown that, in combination with a 3-axis magnetometer (ref paper 13) it could also yield dead reckoning navigation, with potential for combination with either terrain matching or GPS Navstar at lower cost than IN.

In a thoughtful general paper (17), Price then reviewed the impact of Very Large Scale Integrated Circuit technology on the system design process and its costs, and showed how Artificial Intelligence and similar techniques may come to be applied. Surprisingly, the paper, which might have been expected to give rise to debate, provoked none, presumably for lack of experts in the audience. By contrast, the next paper produced more discussion than any other. Basing her argument on projections of expanding requirements for software and increasingly severe shortages of software engineers into the 1990's, Anderson (18) considered that the ADA language alone cannot solve the defence software problems, and advanced the potential of reusable software components as a further aid. High software productivity in Japan was already said to be making considerable use of it, based on cataloguing and retrieval facilities which may be automated or even Expert Systems in themselves; the potential for development along these lines is evident. Discussion ranged over the DOD's plans in the area, the need for standardisation, the differing approaches in the US services to standard ADA environments and compilers, and the problems which arise from the production of requirements and specifications by people who do not understand software technology, giving rise to unreasonable demands for software and increasing the shortage of software engineers.

In one of the few papers not directly aimed at cost reduction, Reilly (19) described a distributed data base management system, aimed at reliable and survivable operation in naval vessels. Though costly, however, it may well be the cheapest way of achieving the high survivability aimed at.

This session was more a collection of interesting and diverse papers than a coherent approach to the session theme.

Session IV – Reduction of Development and Support Costs.

The use of simulation to reduce development and flight test costs has perhaps been more comprehensively practised in

the guided weapon industry than any other for many years. Gauggel (20) reviewed the arguments and experience and described some recent advances relating to infrared AAM development. In the aircraft field simulation is just as valuable and Helie (21) described its use for optimising the pilot-cockpit interface in the context of the rapid advance of electronic display and controls.

Standardisation is a powerful tool in cost reduction efforts and it must have come as a surprise to many not working in the field when Henne (22) revealed the total lack of standardisation (hence total confusion) of the interfaces between aircraft and their many types of weapon; he described the USAF efforts to introduce a standard (MIL-STD-1760) to rationalise this situation.

Session V — Examples of Cost-Effective Accomplishments and Approaches

Terrain following is a very necessary but, in its normal form, very costly technique for attack aircraft penetration. An inexpensive method of achieving it, using a pre-stored flight altitude profile, was described by Schwegler (23). The system had been flight tested at operational speeds, but the crucial details of the ground clearance achieved were not revealed.

In the next paper (24) Brunner described the use of wing-mounted pressure taps as a means of sensing lift coefficient for the control of STOL aircraft at low speed, where control by speed is undesirable. Promising results had been obtained and further work is proposed, including, as suggested in discussion, measurement of sideslip by a fin-mounted sensor.

As demonstrated in several papers in this symposium, a good approach to affordable systems is to integrate sub-systems designed to make full use of information available elsewhere in the system. Paper 25, presented by Towler, described a simplified inertial navigation unit designed to take advantage of information available from a wide variety of other sensors, particularly Doppler or GPS-Navstar. Interest in the discussion centred on the well known Ferranti technique of carrying a temperature model of the system to avoid the need for thermal stabilisation and facilitate rapid alignment.

The final paper (Gloerud, 26) described the integration of the Penguin Mk 3 missile guidance and control system using 6 standard microprocessors connected by a high capacity data bus. It clearly demonstrated the advantages of this approach in reducing development time and cost. The question and answer session ranged over the performance of the alignment system, software design and computer characteristics.

ROUND TABLE DISCUSSIONS

Finally an active general discussion was aimed at crystallising the lessons of the conference. The very wide range of costs of gyros was noted, and it was pointed out that this is not just a matter of design, but for a given type depends strongly on the specification to be met. However system cost depends very much on how the inertial instruments are configured, and there is much scope for ingenuity here. The trend seems to be towards fully integrated systems where the same sensors, together with standard data buses and distributed computers, supply the needs of the entire guidance and control system. Doubts were still expressed, however, concerning the basic conflict of requirements between the flight critical system's need for reliability, simplicity and transparency, and those of, for example, navigation for high accuracy; this may be reflected in high costs in verification and testing. Clearly no one approach is likely to meet all needs.

More and more of the rising costs of defence equipment is being devoted to software, and in general, where both hardware and software solutions are available, software is usually much cheaper. Nevertheless the need to devote attention to reducing software costs is very clear. Standardisation of high level language and data buses are beginning to make an impact, and reusable software should help in the future. Software engineers however are seldom closely aware of the real system needs, and system engineers have a heavy responsibility for cost reduction. Generally speaking, the structure of the procurement process imposes too many agencies between the customer and designer. The initial statement of a requirement, by the nature of things, always involves uncertainties as to what is really needed, and what the cost implications are. There is therefore a need for easy, rapid interaction between customer and designer to achieve the best compromise between performance and cost. Every error, misconception and omission removed from a specification or statement of requirement represents a cost saving. There is however a potential difficulty if customers and designers are to be closely coupled to achieve minimum cost, namely how to reconcile competitive bidding with such a process?

Finally the Symposium Chairman summarised the conclusions of the meeting thus:

- Operational requirements must be very carefully produced;
- Frequent modifications must be resisted unless very strongly justified;
- Co-operation between specifiers and designers is essential at all levels, and especially early in the process;
- More effective collaboration is inevitable, starting with the sharing of development costs;
- Consideration of saleability is always important;
- No matter how smart our systems become, numbers remain important, therefore thinking in projects must replace thinking in financial years;

and high quality engineering is more than ever vital.

4. AUDIENCE REACTION

Interest in the symposium was demonstrated by a consistently high attendance which was maintained to the end. As usual, questionnaires requesting views on the meeting were issued to participants, but only about 1 in 6 responded, leaving doubts as to how far the replies received can be taken as truly representative. Responses to the initial question may be summarised as follows:

| | Good | Average | Poor |
|---|-----------|---------|-----------|
| a1. Quality and relevance of papers, sessions and questions | 9 | 5 | 0 |
| a2. Did the papers support the theme? | 5 | 11 | 0 |
| | More than | Yes | Less than |
| a3. Did the symposium live up to your expectations? | 2 | 9 | 5 |

In answer to the more general questions, there was satisfaction with the technical content, but some disappointment that the user voice was not represented, either in the papers or the audience. The critical needs for R & D were felt to be in improving software engineering, making progress with VLSI/CHSIC, and increasing reliability and maintainability. Major challenges and unresolved problems for the future were stated as all round cost reduction, better standardisation and better statements of user needs and requirements. Asked what further action NATO and AGARD should take in this area, several suggested the promotion of standardisation, while two suggested returning to the topic in 5 years time.

5. TECHNICAL APPRECIATION

The keynote speaker emphasised the importance of using every possible approach to the reduction of costs, and this was reflected in the wide variety of topics addressed in the programme. Nevertheless there were some important omissions. Absence of a user contribution has already been referred to, and there was also a lack of papers dealing comprehensively with cost reduction in the design and development of major systems, a point noted by several of those responding to the questionnaire. Perhaps the paper by Gioerud came nearest to this, and certainly if it is taken together with previous presentations on Penguin (eg Brodersen, Proc 36th GCP Symposium, May 1983) a comprehensive picture does emerge. A number of papers addressed general approaches to the subject, of which those on design-to-cost (Shapiro), reliability improvement (Capitano) and software engineering (Anderson) were notable. The wide range of papers on low cost design in guidance and control, mostly at the sub-system or component level, were generally of a high standard of technical interest. An excellent effort to fill the gap in between was made by Ostgaard, in what may very regrettably have been his farewell appearance, and it is unfortunate that his paper cannot be included in these proceedings. Another paper at this level, that by Murgue and Evrard, provided an object lesson of interaction between user and designer to meet user needs at minimum cost, underlining the necessity of that interaction as a vital contribution to the control and reduction of defence equipment costs.

6. MILITARY POTENTIAL

There can hardly be a topic which should be of greater interest to the military services than the subject of this symposium, and it is astonishing that there was so little military participation. Since some of the most valuable comments relative to the part the user must play if lower costs are to be obtained came in the discussions, and cannot be fully reported here, an important opportunity has been lost. However if the point can be registered that there is a strong feeling in the technical community that a more flexible approach to operational requirements and specifications through interaction between user and designer can make a major contribution to reducing costs, there will be a basis for real progress. In addition, these proceedings can be recommended to the military staffs for their delineation of a wide variety of approaches to the achievement of effective and affordable systems.

7. PRESENTATION AND ADMINISTRATION

Not many conferences can have given rise to so few complaints. The facilities and assistance provided by the host country, Turkey, were excellent, as was the AGARD organisation. The Programme Committee had a difficult task in putting together a coherent programme, given the nature of the topic and the diversity of the papers, but on the whole they succeeded. It was evident that attention had been paid to persuading contributors to speak clearly and not too fast, and there were very few who caused the interpreter any difficulty. Visual aids were generally good, and a number of the French speakers had thoughtfully provided slides with English or bilingual captions, a courtesy which those speaking in English should consider returning in the future. The conference had the great advantage of having all participants in a single hotel, so facilitating technical and social interaction throughout.

8. RECOMMENDATIONS

The Guidance and Control Panel is to be congratulated on having successfully tackled a subject which many believe, notwithstanding its importance, might be too difficult to form the basis for a successful symposium. Undoubtedly, what was

progress is made in the future, the problem of the high cost of military systems will not go away, and it is therefore recommended that a further symposium on the subject should be considered in about 5 years.

It may also be noted that a number of participants in the conference recommended increased NATO effort towards standardisation, though this is not a field in which AGARD could be expected to play a major role.

Finally it is recommended that steps be taken to acquaint NATO military staffs with the conclusions of the symposium.

**REVIEW OF MISSION VERSUS COST REQUIREMENTS
AND TECHNICAL ISSUES**

by

C. Baron
Head, CM(s) 1/SAS
Room 221 Ministry of Defence
Savoy Hill House, Savoy Hill
London, UK

Very probably everyone in this audience would agree that we have a problem in the rising cost of aircraft, and, of more direct concern to us, of guidance and control systems; in the first part of this talk I will be discussing this problem in general terms, drawing heavily upon a paper by Kirkpatrick and Pugh (Ref 1). I hope I shall succeed in convincing at least some of you that the problem is more serious even than you think. I shall then briefly refer to a couple of examples of weapon system developments in which I have been personally involved, and try to draw lessons from those particular instances which bear on the subject. Then finally I shall come back to general principles and point to a route which, at least in some instances, may get us out of our basic dilemma, that of seeking more fighting capability but not being able to pay more for it.

The rise in combat aircraft costs is not just a recent phenomenon, as many people think; Fig 1 shows that it has been going on almost since the start of military aviation; this refers mainly to British aircraft, but the inclusion of a number of US types shows that the trend is basically the same on both sides of the Atlantic. The diagram plots unit production costs, compensated for inflation, as a function of time. Cost escalation is not just a thing that applies to aircraft, or a few similar types of system. As this table from Sy Deitchman's excellent book (Ref 2) shows, it happens over a wide range of weapons and systems, and it seems reasonable to assume that very little operational military equipment is not involved in inexorably rising costs.

ANNUAL REAL GROWTH OF UNIT COST, PER CENT

| | |
|------------------|----|
| Infantry AT | 13 |
| Tank | 11 |
| Destroyer | 9 |
| Aircraft | 8 |
| Aircraft Carrier | 6 |

There is of course a natural human tendency to assume that it is the other fellow who has the problem, not ourselves, and I have spoken in the past as much as most people I suppose about the rapidly reducing cost of solid state electronics and computing, always foreseeing a more advanced and even cheaper stage ahead. Fig 2 shows what has really been happening; we see that for both all weather and VFR aircraft avionics the annual cost rise is about the same as that for aircraft as a whole. Why is it that we have not done better than this, in view of the advantages we have had? The answer, I think is that we have been encouraged to use, for example, the spectacular reductions in computer hardware costs to seek improvements in performance, not cost reductions; and in doing so we have developed needs for better or additional sensors to feed them, for better or additional actuators etc to use their increased outputs, and of course for mountains of software.

Kirkpatrick & Pugh have pointed out that this phenomenon of rising defence equipment cost arises from the operation of positive feedback in the procurement process. Fig 3 depicts the first and most obvious example - when two nations or power blocs see each other as potential enemies, each new aircraft on one side, Red, say, is seen as a threat by the other, forcing a technology advance to meet this threat which is seen as being so vital that the cost of achieving it becomes a secondary, though still important factor. The resulting new Blue aircraft stimulate the same process on the Red side and so a positive feedback loop is established. It is important, I think, to point out that although in this process one side or the other will get ahead for short periods, the average over time is inevitably a drawn game, and the only steady trend with time is a continual increase in cost to both sides.

Fig 4 shows four more vicious circles - on the production side, higher costs lead to smaller numbers being produced, hence lower investment, less learning and so ever higher production costs. Similarly in the development process, higher costs mean less frequent projects, so bigger jumps in technology are needed. Less frequent projects lead to decay of development teams between projects with the result that they are less able to cope with the advances in technology required, and development costs and times rise ever more steeply. And all this makes decision-taking more difficult, each decision gives rise to more and more studies, investigations and argument, and the process becomes ever more protracted and ever more costly.

The problem of cost escalation is so fundamental and so severe that every possible approach which may oppose it must be employed. One important contributor must be the

exercise of great care in laying down performance requirements. As Fig 5 shows, when a new type of system becomes available, rapid improvements in performance are achieved at first as more and more ideas and research are applied to it. After some decades, however, more fundamental limits to performance are reached, as for example, the limitations of materials and structural technologies are tending to set limits to the speed of aircraft at low altitude. If now a performance is called for which is too close to the limits of technology, large cost increases will arise, or the project may even have to be cancelled. The problem of the Service in choosing the right level of performance to specify is exacerbated by their having been accustomed, for decades perhaps, to obtaining major performance rises with each new generation - moreover, in complex systems it is not always clear just where the limits will be.

To give you an example from my own experience, back in the 1960s, when high speed low level aircraft began to be perceived as a threat, both the US and British armies produced requirements for low level SAM systems, and developments were begun on both sides of the Atlantic. The requirements were much too ambitious, however, and the British system, called, PT428, was cancelled after two years, and the US Mauler after four, in both cases because of rapidly rising costs and serious development problems. Now I have to admit that I, along with other engineers and scientists, had gone along with the Armies' requirements without protest, but at least with the cancellation of PT428 I did, with a colleague, Mr J E Twinn, begin to think out what the real need of the Army was, and we were able to evolve a proposal for a system which could provide an effective low level defence with a unit cost almost a tenth of what PT428 and Mauler had been expected to cost, and with lower operating manpower. We had great difficulty in persuading the Army to accept our proposal, since it meant radically changing their existing policies and performance requirements. Fortunately we succeeded in the end, and our proposal became the Rapier system (Fig 6), which has been sold to the tune of over 3 billion pounds in many countries, and for the invention of which, as the Chairman kindly mentioned in his introduction, Mr Twinn and I recently received a major MOD award.

Now the lessons I would like to draw from this experience are, first, that technologists must play an active part in helping the Services to pitch their requirements so as to yield equipments with the best compromise between performance and cost, and secondly, that equipments designed for the lowest cost in conjunction with acceptable performance can still be technically interesting. Rapier was a more original and innovative design than either PT428 or Mauler, because we had a stronger incentive to introduce new ideas to keep the cost and manning down.

For example we designed an unmanned pulse doppler radar which, at relatively low cost, automatically performed the functions of target detection and location, threat assessment and identification, none of which had ever been attempted before.

For any equipment at a given stage of technology, the relationship between cost and effectiveness is of the form shown in the upper curve of Fig 7. At the lowest cost we have very simple equipments which are virtually ineffective, but then we reach a threshold where effectiveness increases more rapidly than cost up to the point of optimum unit cost effectiveness, beyond which the approach to the limit of available technology causes rapid cost escalation. If we now consider the effectiveness of a force of equipments, there are two reasons why, for a constant total expenditure, it is better to go for a cheaper equipment than that which gives optimum unit cost effectiveness. This is because of the advantage of having greater numbers, both in reducing unit production costs and in increasing force effectiveness.

One can well sympathise with the problem the Services face in pitching their requirements so as to hit this maximum - if they call for too much performance, even at the level of optimum unit cost effectiveness, they will get only a small number of expensive aircraft. On the other hand, if they go all out for low unit cost they may get a lot of cheap aircraft but their force is equally ineffective.

I remember one very clear instance of this, again from the time when I was much involved with guided weapons system design. We were considering the design of a new naval air defence system, and the Navy, having become very concerned with rising costs, had decided to deal with this by specifying maximum cost figures for both weapon and system. An argument developed about the threat the system should be designed to meet; the scientists said that since it was by then well within the state of the art to design a sea-skimming anti-ship missile, it was essential that the system should perform against such a threat. The Navy, believing that this would lead to a design above their cost limits, and being able to say that there was no sign of such a threat materialising, refused to accept this and the system design went ahead without that capability. The result was that when the threat did appear a few years later very expensive modifications had to be made to the system and the final cost was much greater than it need have been.

Apart from this question of getting the requirements right, other methods of reducing costs include international collaboration, export promotion, improved production technology, and value engineering. In the aircraft field essentially all the benefit likely to be achievable from the first two has already been obtained, and while undoubtedly the application of computers to production will continue to give benefits, the small production runs dictated by modern aircraft costs prevent the application of the capital investment needed fully to apply such techniques.

Value engineering has already produced some substantial savings but may now have reached the stage where further improvements will be smaller and more difficult to obtain. We have to remember moreover that the long history of military aircraft has always seen the constant incorporation of improved production and cost reduction techniques and yet that 8% per annum cost escalation has gone on without check. There therefore seems no clear reason to believe that it will not continue. At best perhaps we may be able to slow it down; certainly we must put every possible effort into doing so. Perhaps the greatest hope is the seriousness of the situation; most nations have now been forced to reduce the size of their aircraft fleets, and the number of types of aircraft they employ, to a minimum below which the viability of the air force comes into question, and even holding the force at this level has required transfers from other areas within defence budgets. Moreover the high costs of the now infrequent aircraft purchases throw serious strains on defence budgets when they do occur; at its peak Tornado absorbed 18% of the UK defence budget (Ref 3) and similar problems have been observable in most Western European nations in recent years.

One possible way out of this situation may exist. Consider again the way in which system performance increases with time as depicted in Fig 5. When a new type of system is introduced, be it aircraft, tank, torpedo or whatever, improvements at first come rapidly and cheaply; after a few successive generations, however, a law of diminishing returns sets in, progress is only obtained slowly and at great expense; at this stage it also becomes increasingly difficult to achieve a decisive technological advantage over an opponent, because both sides tend to be pressing closely against the same fundamental limits. My ex-chief, lately Chief Scientific Adviser Sir Ronald Mason has graphically described this situation as that of 'baroque technology', as opposed to new technology when we are operating in the lower part of the curve. Suppose we could replace our baroque technologies by new technologies, we would once again be in a situation to move forward rapidly and cheaply; moreover we would be able to move away from the naturally drawn game of the baroque technologies to one in which we would have a lead, enabling us to dictate to the opposition and possibly impose upon him much greater costs than we need expend ourselves.

This has of course happened frequently in the past in the replacement of sail by steam, the horse by the tank, and so on. Consider the development by the Royal Air Force of high speed low level aircraft attack techniques; it cost us a good deal in the development of aircraft, navigation, cockpit instrumentation and weapons, but when one considers the enormous efforts the Soviet Union has had to put into low level air defence, only recently culminating in AEW and look down shoot down interception systems, there can be little doubt where the advantage has lain. Now would be about the right moment to change again to a new technology, and it may be that we have it in the shape of the cruise missile and similar systems, which have the characteristics to negate the opposition's previous air defence efforts, and send him back to the drawing board again.

It is important to recognise that new technology can only yield these advantages provided it is allowed to replace baroque technology. One has seen, for example satellite communication systems being introduced to supplement, but not replace, the previous communications systems; new navigation systems are all too often added to the aircraft cockpit, yet another complication for the over-worked pilot to master, while none of the old types are removed, in the belief, probably mistaken, that more equipment will yield better navigation.

There are at the present time great possibilities open to us in the applications of new sensors, advanced signal processing and artificial intelligence, to name a few outstanding examples. If we merely employ these to add to the equipment of baroque system types we will only contribute to further cost rises. If, on the other hand, we are able to apply them to introduce new types of systems which make possible new approaches to military problems we can not only reduce our own defence costs, but at the same time, by taking the initiative from the opposition we may also increase the strain on his resources. The scientific and technical community must work to counter the forces of conservatism and vested interest which always tend to oppose new ideas.

One last point: the Western nations have all tended, to a greater or lesser degree, to seek to employ sophisticated equipment to allow them to reduce the number of people in the Armed Service, conscription being unpopular in democracies. The UK in fact, has done this more than most. The Eastern Bloc, on the other hand, has little difficulty in keeping in being very large forces. If the cost of manpower increases in line with inflation, while equipment costs rise 8-10% faster, this Western policy must become untenable in the end.

It is therefore absolutely vital that we should all play our part in the effort to reduce equipment costs, and to see that technology is applied to reducing costs, and is not allowed to become a reason for increasing them.

References

1. D L I Kirkpatrick and P G Pugh - Towards the Starship Enterprise - are the current trends in defence unit costs inexorable?
Aerospace, May 1983.
2. Seymour J Deitchman - New technology and military power
Westview Press, 1979.
3. HMSO, UK - Statement on the Defence Estimates, 1982.

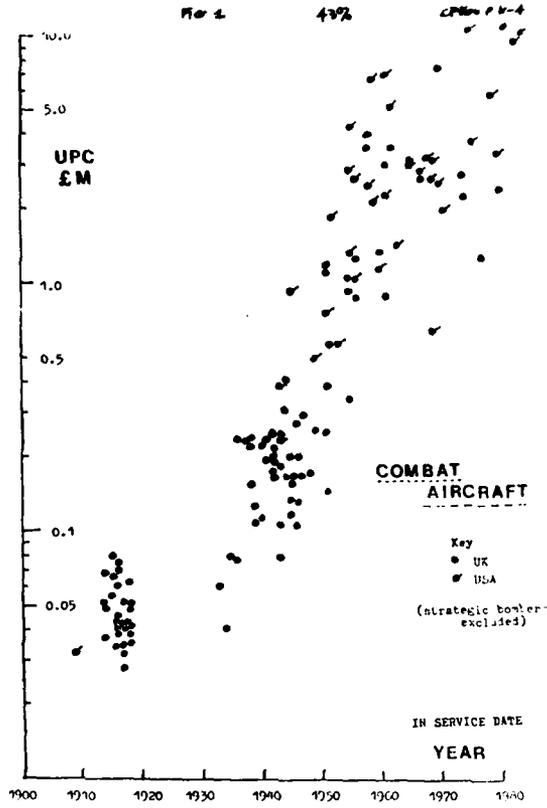


Figure 1 The historical perspective

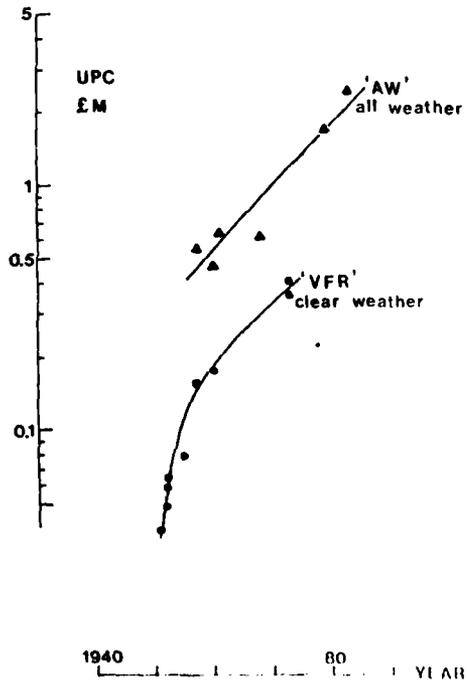


Figure 2 The rising real cost of mission avionics

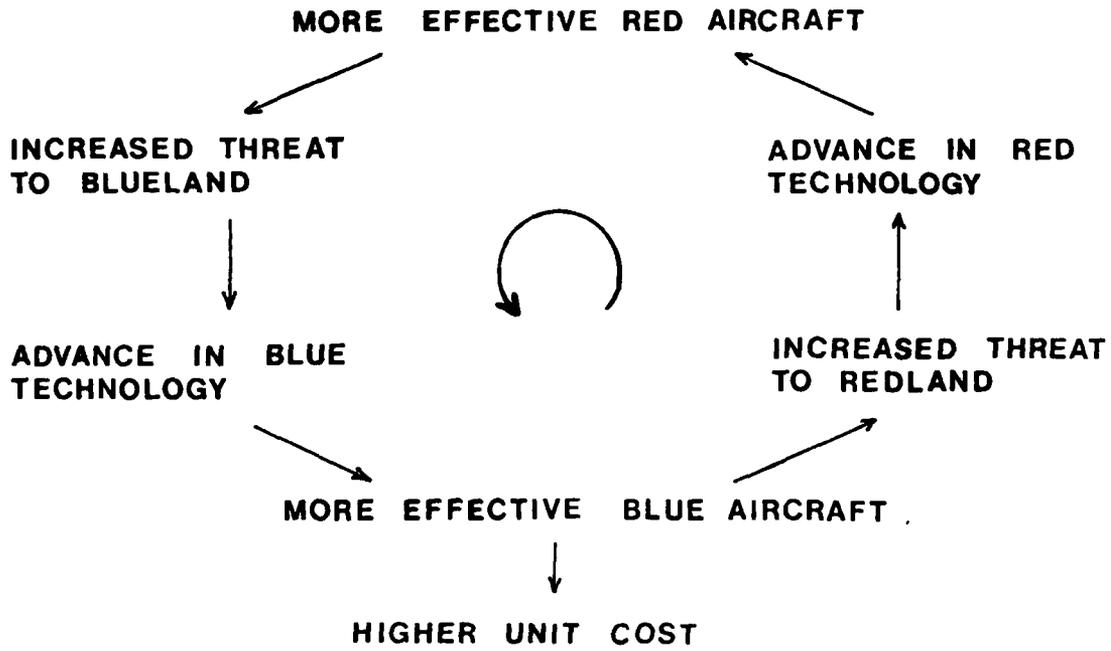


Figure 3 Positive feedback causing increasing aircraft cost

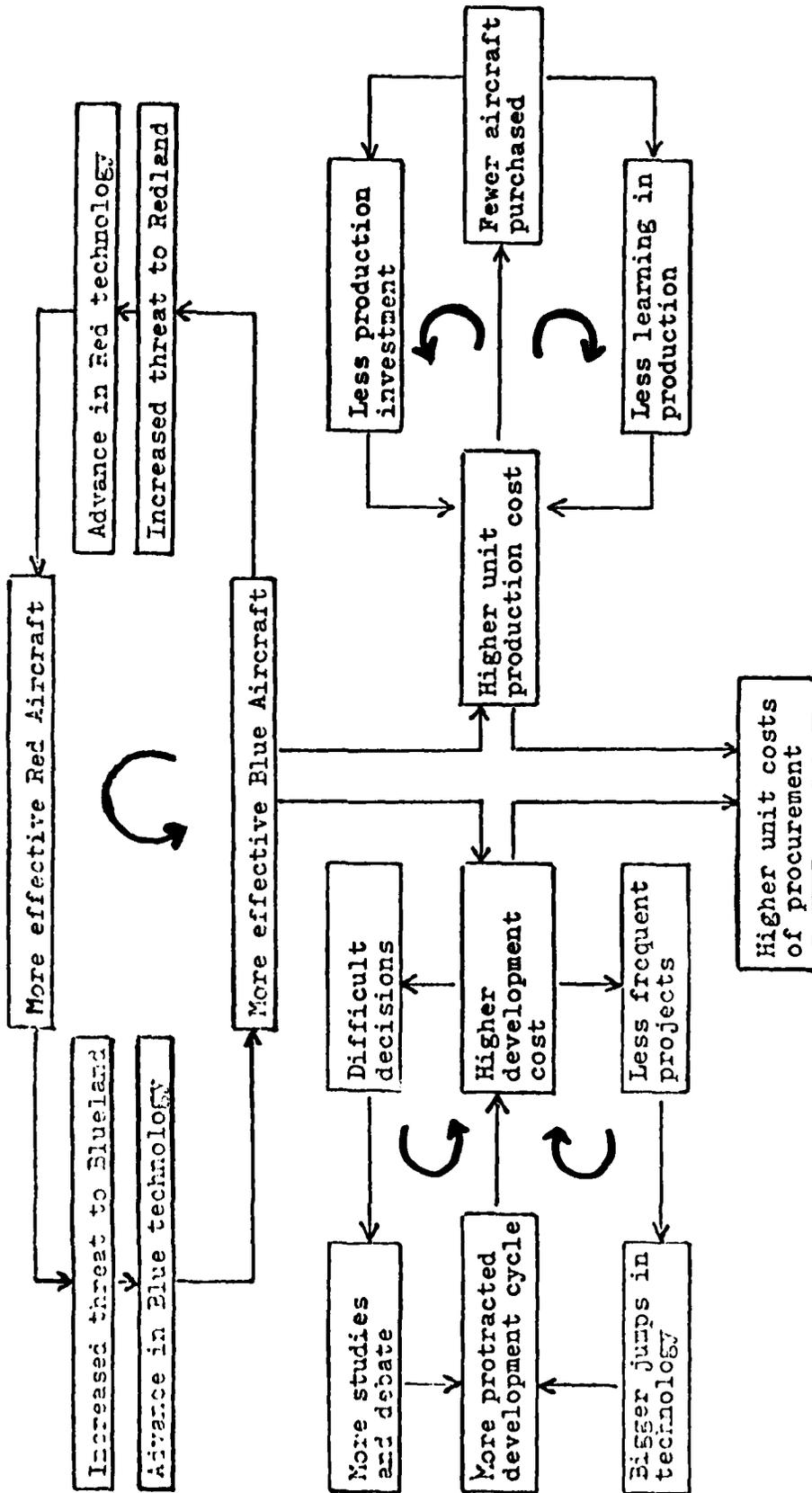


Figure 4 Vicious circles

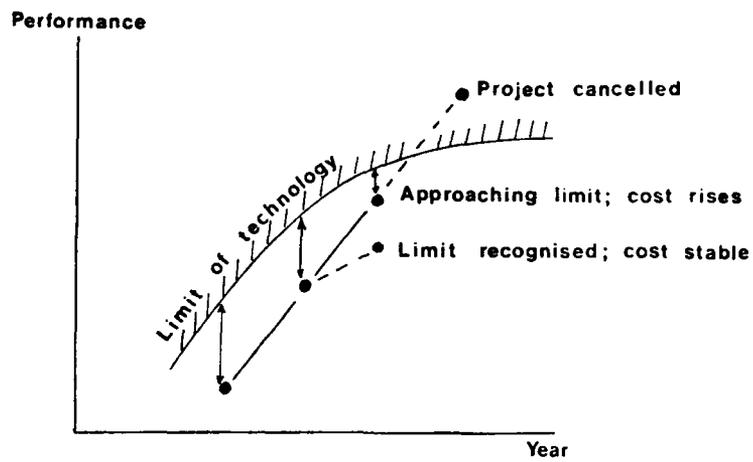


Figure 5 The danger of specification by extrapolation



Figure 6 The Rapier ground-to-air missile defence system

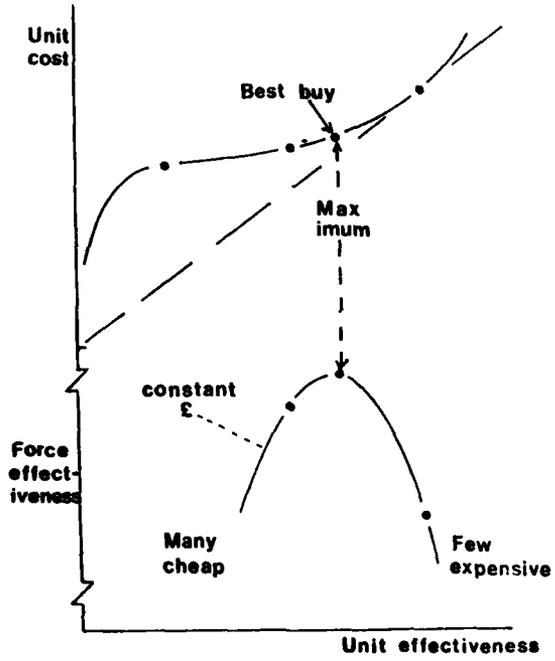


Figure 7 Maximising force effectiveness

SOME ASPECTS OF HOW TO DESIGN COST-EFFECTIVE
FLIGHT CONTROL SYSTEMS

by
U. Butter and L. Botzler
MESSERSCHMITT-BÜLKOW-BLOHM GMBH
Aircraft Division
8000 München 80, Postfach 801160
FRG

SUMMARY

The design of flight control systems for fighter aircraft is discussed with respect to areas which contribute to minimizing life-cycle costs. As life-cycle costs include all costs accumulating during the whole life of the system, all phases from the design to in-service use are considered. Any structural and technological design features that are introduced to save costs during system operation and maintenance require additional development effort. Therefore, the expected cost benefit has to be balanced against the development effort invested into the system to achieve a cost-effective design.

The paper is comprised of two chapters, with chapter 1 covering the sensor information processing and chapter 2 covering the actuation system.

1. SENSOR INFORMATION PROCESSING

1.1 INTRODUCTION

Flight control systems are designed to fulfill given performance and safety requirements and to comply with aircraft specific boundary conditions. Apart from that, it is the task of the design engineer to select from the various possible system configurations and available technologies the most cost-effective solution.

Cost-effectiveness means: Minimal life-cycle costs at a given level of performance where life-cycle costs include all costs accumulating during:

- ° the design and development phase,
- ° the production phase,
- ° the in-service phase, and
- ° the post design and system upgrade phase.

It should be noted that the performance and safety requirements have a strong cost impact and should therefore be kept at a reasonable level. In the following, however, performance and safety requirements are considered as fixed values which are not further discussed.

The design features of the system affect the costs in all phases in different ways. Thus, it is necessary that the costs produced in the various phases are not considered in isolation but as a whole in order to achieve an optimal design with respect to minimal life-cycle costs.

1.2 COST FACTORS

The parameters which influence the production costs per system and the associated aircraft costs are system size, weight, power consumption and environmental condition requirements.

The costs, accumulating during the in-service phase, are determined by system reliability, maintainability and testability.

The flexibility needed for modifications are given by system modularity, standardization and growth potential.

The development costs are a function of the complexity of the selected solution, which depends on the above mentioned features that are designed into the system. Apart from that, there are computerized methods and tools that support the design and thereby help to reduce development costs.

1.3 DESIGN FEATURES

Integration of flight control with fire control and propulsion control is supported by the use of multiplex data busses for communication between the flight control system and the avionics and utility systems. Also, data transmission within the flight control system via a multiplex data bus offers a number of advantages:

- ° Mil Std 1553 B, for example, is a worldwide introduced standard. Standard test equipment is available so that no special interface and no special test equipment must be developed.
- ° A data bus, e. g., to connect the flight control computer with an actuator control

unit, has considerably less weight than conventional analog data links and thereby helps to reduce total system mass.

- Mil Std 1553 B is a command-response-system with high reliability and testability.
- Standardized interface makes it easy to exchange existing equipment against a new type of equipment in the course of system upgrade and modernization.

In Fig. 1.1 the structures of a conventional and a data bus oriented flight control system are compared against each other.

Further to system structure, Fig. 1.2 a shows a flight control system which consists of three separate subsystems (autopilot, command and stability augmentation system, spin prevention and incidence limiting system). By integrating the functions of the three subsystems into one set of computers (Fig. 1.2 b) it would be possible to significantly reduce size, weight and power consumption. Estimated values, taking the present system as a reference, are given in the table below:

| System Structure | Size | Weight | Power |
|---------------------|------|--------|-------|
| Conventional System | 1 | 1 | 1 |
| Integrated System | 0,56 | 0,60 | 0,81 |

Beside that, 5 different Line Replaceable Units (LRU's) would be replaced by 3 LRU's of the same type. This would require less spare parts.

Not always is system integration so straight forward with regard to weight reduction, as is shown with the following example.

A study has been carried out to compare three different configurations of actuator control as part of the flight control system (Fig. 1.3):

- (a) Actuator control units installed in the rear of the aircraft fuselage with data transmission between flight control computers and actuator control units via a Mil Std 1553 B data bus.
- (b) Actuator control integrated into flight control computers.
- (c) Actuator control units and flight control computers both located in the avionics bay.

Estimated values of size, weight, and power consumption, taking solution (a) as a reference, are shown in the table below:

| Solution | Size | Weight | Power |
|----------|------|--------|-------|
| a | 1 | 1 | 1 |
| b | 0,75 | 1,09 | 0,81 |
| c | 0,88 | 1,12 | 1,02 |

Solution (a) turns out to have the lowest weight. In this case, using a data bus saves more weight than would be achieved by higher system integration. However, the reliability of the actuator control units is reduced, because they are installed in an unconditioned area of the aircraft. If for the actuator control units special cooling is provided or high temperature resistant electronics is used, the development and production costs go up. With respect to size and power consumption, the integrated solution (b) is the best.

Generally speaking, system integration offers:

- Reduction of size, weight and power consumption.
- Better testability.

On the other hand, a modular structure offers:

- Less expensive spare parts
- Changing LRU's is easier due to smaller weight.
- Higher flexibility for modifications.

Inertial sensors are required for flight control, navigation, and weapon delivery. However, flight control requires body angular rates and translational accelerations from redundant sensors, whilst navigation requires position and velocity data with high accuracy. Therefore, present aircraft have rate gyros and accelerometers for flight control and inertial navigation systems for navigation and weapon delivery (Fig. 1.1 a).

Strap-down inertial sensors with digital data processing provide both, body rates and accelerations for flight control and sufficient accuracy for navigation. So, these multi-function sensors can be used to supply the flight control system and the avionics systems with the necessary sensor signals, thereby reducing weight and volume (Fig. 1.1 b).

Air data are required for flight control, air intake control, propulsion control, and cockpit instrumentation (Fig. 1.1 a). An integrated air data system, which supplies all aircraft systems with the necessary data, again helps to reduce weight and size (Fig. 1.1 b).

The introduction of advanced electronics technology offers a number of advantages. A reduction of power consumption is achieved with:

- VLSI standard processors and peripheral components,
- Programmable Array Logic and C-MOS Gate Arrays,
- Fiber optical links.

In particular, C-MOS Gate Arrays offer:

- Less components and soldering points because of high complexity circuits.
- Less power, less cooling air and improved reliability because of low power dissipation,
- Higher clock frequency,
- Reduced development costs by application of Computer Aided Design (CAD),
- Reduced modification costs.

The leadless chip carrier technique on ceramic multilayer cards offers:

- Improved reliability due to small temperature differences because of excellent heat conduction,
- Higher clock frequency,
- Less space required compared to conventional Dual In-Line Package (DIP) components.

Standardization of interfaces (e. g. Mil Std 1553 B), modules (e. g. power supply units), and components (e. g. receivers) helps to simplify system modification.

The computing power (throughput, memory) and the interfaces should have a certain percentage of growth potential to allow for small modifications without a complete hardware redesign.

Self-test, i. e. continuous and interruptive built-in test, has the task to detect and isolate failures prior to and during flight to ensure flight safety and to identify and locate failures to support maintenance.

To be operationally cost-effective, the pre-flight test shall fulfill the following requirements:

- Short duration of test,
- Minimum pilot's participation during test,
- Minimum movement of control surfaces during test,
- Avoid nuisance No-Go indications.

Failures identified by the continuous monitoring shall be either indicated on a maintenance panel or recorded on a maintenance recorder. Only relevant failures shall be stored to avoid unnecessary maintenance activity.

The first line test on the aircraft shall be capable of identifying and locating failures down to card/module level rather than LRU level. The advantage of this is that, after the defective LRU has been removed from the aircraft, there is no additional test necessary to identify the defective module. This reduces test time and test equipment.

The equipment shall be installed in the aircraft in such a way that the maintenance crew has easy access to it for inspection and LRU replacement during turn-around.

Sufficient burn-in shall be applied to have a good selection of components and achieve high reliability. Improved reliability means reduced defect rate, higher equipment availability and cost savings because of:

- Less frequent testing for identification and location of defects and proving of serviceability after repair,
- Less frequent changing and repair of defective equipment.
- Less spares required.

1.4 DESIGN AND TEST TOOLS

The increasing requirements on flight control and the progressing integration with other aircraft systems make flight control systems for modern fighter aircraft more and more complex. This requires that computerized methods and tools are used to support the design, test, documentation, configuration control, and project management in order to minimize design errors and reduce development time and costs.

Examples of design methods are:

- Structured Analysis
- Modular Design
- Pseudo-Code

Structured Analysis supports the design by means of a complete and consistent system model. Modular Design cuts the system down to small and transparent units. Pseudo-Code provides an easy-to-understand description of the system functions.

Software integrity is a pre-requisite for flight release of a digital flight control system. One possibility for verification and validation of flight control software is to use another computer with dissimilar software for a comparison test. This method has successfully been used to test the software of a digital autopilot. For this purpose a Cross-Software Test-System was developed with a test computer, that contains a software model of the autopilot as well as a stimulation and evaluation program. The test is fully automated.

2. HYDRAULIC ACTUATION SYSTEM

2.1 LIFE CYCLE COSTS

The life cycle cost shown in Fig. 2.1A is comprised of:

- ° Development Costs
- ° Manufacturing Costs
- ° Maintenance and Operational Costs

The development cost constitutes approximately 10 %, the manufacturing cost 20 % and the maintenance and operational cost 70 % of the total life cycle cost. This makes it understandable that an attempt should be made to invest more in development and less in maintenance and operational costs whenever possible.

To this end, the following points should be kept well in mind:

- ° Sophisticated, straight-forward design philosophy involving simple solutions (no elaborate hardware).
- ° Modular design, the modularity striving toward common building blocks for the various actuation systems.
- ° Easy access to components to facilitate maintenance.
- ° Common actuation systems for the various control surfaces whenever possible.
- ° Automatic failure diagnosis using built-in test (BITE).
- ° Extension of life time by replacing modules only upon a condition involving defects (conditional maintenance).
- ° Easy alignment procedure for remounting overhauled equipment.

A further aspect increasing the life cycle cost involve indirect costs which may be reduced by:

- ° Constructive designs that minimize wear, or fatigue behaviour.
- ° Automatic test procedures (BITE) that do not evoke unnecessary wear.
- ° Minimizing the scope of inspection procedures for periodic maintenance.
- ° Low commitment of manpower in the maintenance effort.

To give an overall view, Fig. 2.1B shows how the life-cycle cost of maintenance and operation is distributed throughout all equipments in a fighter aircraft. 50 % of all costs involve the flight control and avionics systems.

2.2 FLIGHT SAFETY COSTS

Flight safety can only be achieved by high levels of redundancy which includes multiple channels, identical in technological design and having static and dynamic characteristics within close tolerances to achieve low monitoring thresholds. The close tolerances drive the costs upward; therefore, specified tolerances must not be specified any closer than absolutely necessary to ensure flight safety with respect to the transient behaviour of the aircraft after channel failures.

Simplified dissimilar backup channels should be envisaged and allowance made for degraded performance. The failure philosophy should establish "essential" and "non-essential" control surface combinations. The essential surface combination provides a fly-home capability and has a fail-operate/fail-operate failure characteristic. The non-essential surfaces have fail-safe only.

The flight control system must be capable of being "reset" by the pilot after ostensible failures in order to counteract nuisance monitor trips which never can be completely avoided in a cost effective design.

2.3 MISSION RELIABILITY COSTS

Reducing the costs of flight safety influences the mission reliability. Mission reliability should not be overemphasized since many missions can be completed without full aircraft performance capability. The cost of an aborted mission should be carefully weighed against the costs of increasing the redundancy level to maintain full performance of the flight control system after failures. It is the opinion here that flight safety combined in a fly-home capability provides acceptable mission reliability. No extra costs should be accepted in this area with respect to flight control.

2.4 COMBINATION-CONTROL-SURFACE CONCEPT

A combination-control-surface concept divides the control surfaces into essential and non-essential groups. The essential group of surfaces is necessary to provide a fly-home capability after failure in one of the non-essential surfaces. As already mentioned above, the essential surfaces have fail-operate/fail-operate failure characteristics while the non-essential ones have fail-safe only. Fail-safe is accomplished by switching the actuator to heavily damped ("slugged") operation whereby the chambers on each side of the actuator piston are connected together via a restricted oil passage. A combination-control-surface concept is illustrated in Fig. 2.4 A.

The simplexed non-essential control surfaces offer large savings with respect to cost and weight for both the actuation as well as the hydraulic-supply system. Only the mission reliability may decrease as a result of these savings. However, repeating what has been said, many missions do not require high performance in maneuvering in order to be completed.

2.5 REDUNDANCY CONCEPT

Fig. 2.5A shows a redundancy concept comprised of essential and non-essential surfaces. The essential surfaces have quadruplex technology for the actuation systems and associated signal processing. The non-essential surfaces have duplex technology to meet the safety requirements mentioned in 2.2. As an option, a mechanical backup is drawn in dashed lines to the essential surfaces in the event that dissimilar redundancy is considered prudent for possible total failure of the quadruplexed sensor information processing system.

The rudder control is not considered to be absolutely essential for crosswind landings up to moderate wind velocities and therefore the backup option is omitted. The touchdown maneuver may be somewhat rough, but the aircraft can be successfully landed under emergency conditions.

Fig. 2.5B shows a cost-effective redundancy arrangement for a hydraulic supply system. The system is designed to give fail-operate/fail-operate characteristics for feeding the essential control surface actuators. Two of the three hydraulic circuits supplied by each of the two pumps are used to supply the flight control surface actuators. The non-essential control surface actuators are connected such as to balance flow rates as far as possible.

An auxiliary drive is shown as an option. With regard to installing auxiliary power drives, it is the opinion here that the probability of a double flame-out of the engines must be carefully weighed against the costs and loss of aircraft performance through weight penalties.

Fig. 2.5C shows an installation schematic of the supply system depicted in Fig. 2.5B. It should be noted that the long hydraulic lines (most exposed to failure) supply the non-essential surfaces. These lines need only to be simplexed.

2.6 ACTUATOR SYSTEM DESIGN

Fig. 2.6A shows how criteria from aerodynamical and flight mechanical considerations in addition to the required reliability and safety are introduced into the design of an actuation system.

Fig. 2.6B shows a schematic system diagram of a two-servo valve actuation system. The two servo valves work independently upon the control valve. The control valve assumes a "force-fighting" position from the two servo-valve commands and controls the hydraulic flow rate to the main ram (actuator). The feedback signals for the actuation system are the control valve position (ram rate) used for damping and the main ram position itself. These positions modulate a carrier frequency (a few kHz) and after demodulation and A/D conversion, the signals are processed in the sensor information processing system associated with the control surface actuation for a cost-effective design.

Fig. 2.6C shows a drawing of the above described actuation system excluding the information processing. Also, the fail-safe bypass valve for "slugged" operation in the fail-safe condition is not required for essential surfaces. Failure detection in the monitoring system is accomplished by detecting excessive transient excursions from a software modeled servo system. After a failure in one of the servo-valve control loops, the associated 1st-stage bypass valve opens, thus isolating the defective servo-valve. One half of the control valve drive is thereby disabled (bypass operation) while the other half continues to operate. The overall performance of the actuation system is hardly affected.

Fig. 2.6D shows a detailed drawing of the servo valves depicted in Figs. 2.6B and 2.6C and the interface to the information processing system. The mounting of the servo valves to the ram assembly is also shown. Upon failure of one of the four signal channels, the valve continues to operate since the electromagnetic force summation of the remaining three channels keeps the valve performance within the tolerance level of the monitoring system.

Fig. 2.6E shows a cost-effective variant of the actuation system described. Here, the hydraulic intermediate stage driving the control valve has been eliminated. Thus, the associated servo-valves which sum relatively low electro-magnetic forces to drive the intermediate stage have also been eliminated.

This direct-drive servo actuator design assumes that the manufacturing tolerances of the control valve are such that jamming is highly improbable. Also, the electronic driving stages now envisaged produce sufficient magnetic forces in the driving coils to guarantee safe operation.

Fig. 2.6F shows an actuator design from NATIONAL WATER using a direct-drive control valve.

3. CONCLUSIONS

As was shown above, the various design features influence the costs in a positive or negative sense during the different phases of the system life-cycle. So, a cost-effective design can only be achieved with trade-off studies.

The size, weight and power consumption of a system can be reduced by using advanced technology components. The benefit gained from this has to be balanced against the higher price for these components.

There is an interactive relationship between the development phase and the operational phase. Increased development effort to improve reliability, maintainability and testability of the system helps to save costs during the in-service phase. Experience indicates, that the costs invested into development, production and maintenance of a system are related to each other as 1 : 3 : 10. If it were possible, for example, by increasing the development effort by 20 % to reduce maintenance costs by the same percentage, absolute life-cycle costs would significantly be cut down. This indicates, in which direction we have to move in order to get closer towards a cost-effective design.

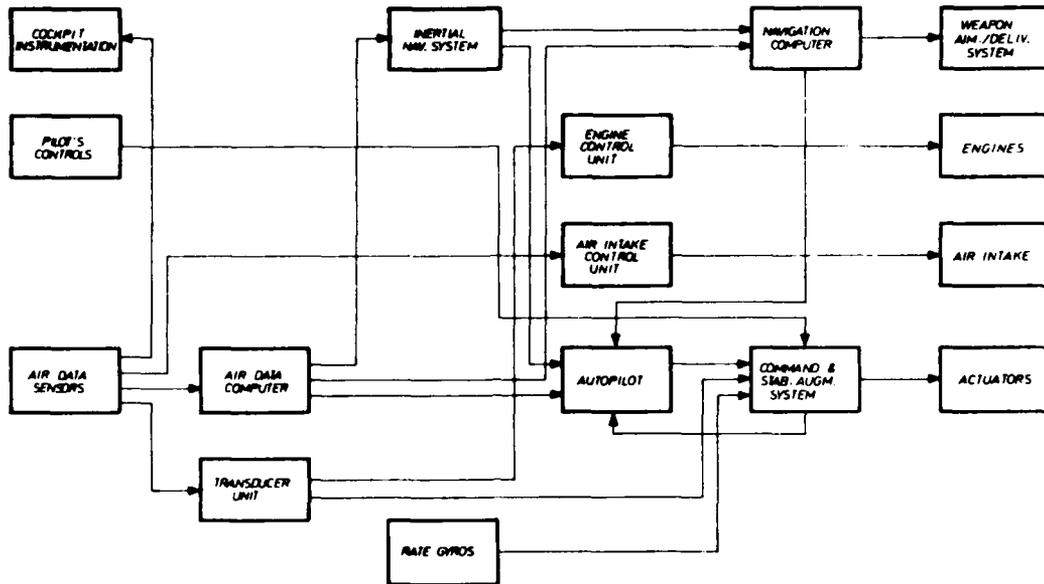


FIG. 1.1A: CONVENTIONAL FLIGHT CONTROL SYSTEM

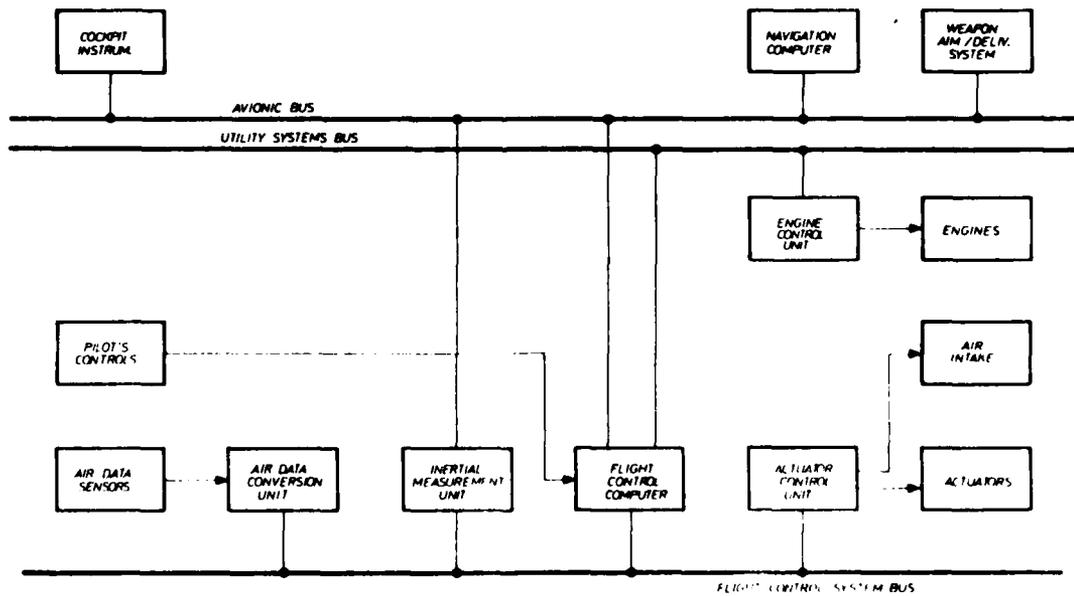


FIG. 1.1B DATA BUS ORIENTED FLIGHT CONTROL SYSTEM WITH MULTIFUNCTION INERTIAL SENSOR SYSTEM AND INTEGRATED AIR DATA SYSTEM

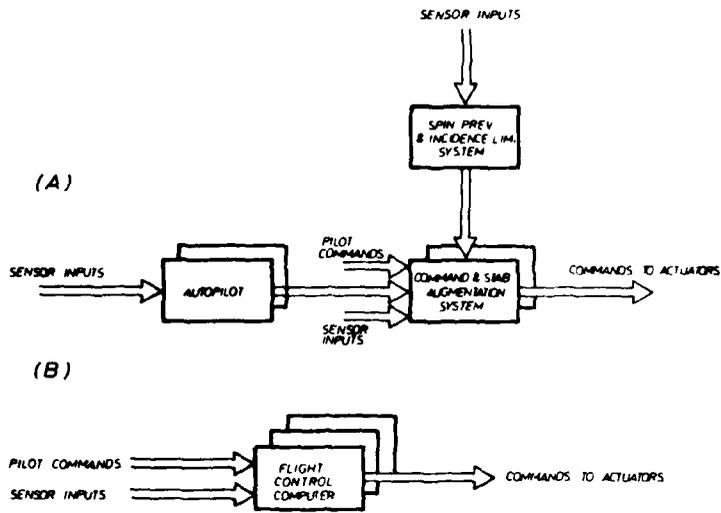


FIG. 12 FLIGHT CONTROL SYSTEM

(A) SEPARATE SUBSYSTEMS
(B) INTEGRATED SYSTEM

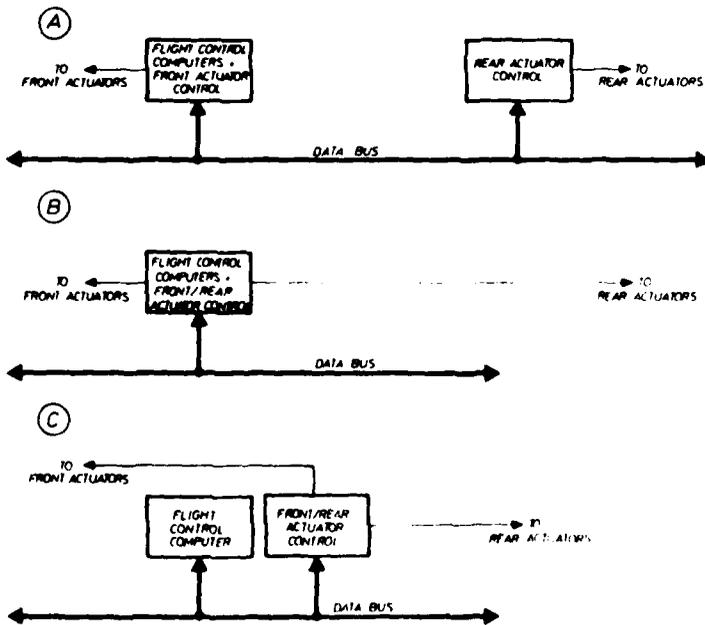


FIG. 13 FLIGHT CONTROL SYSTEM WITH THREE DIFFERENT CONFIGURATIONS OF ACTUATOR CONTROL

COST BREAKDOWN

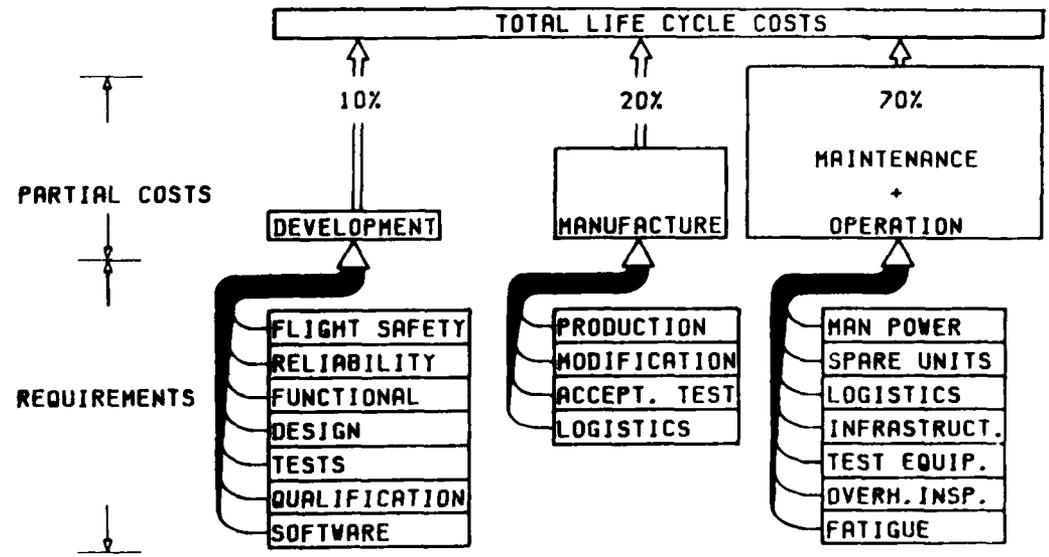


FIG. 2.1A

MAINTENANCE - AND OPERATIONAL COST BREAKDOWN

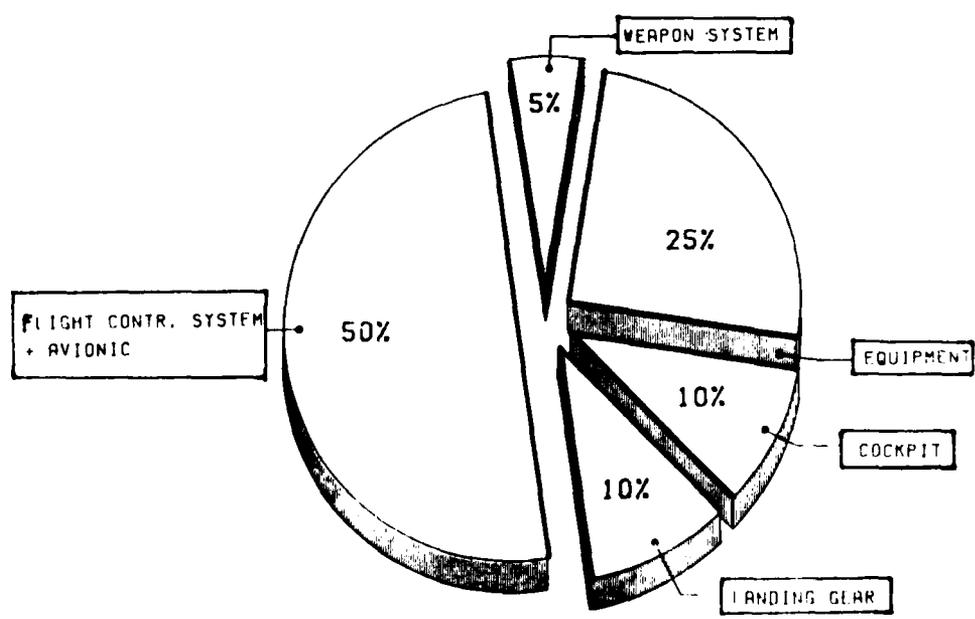


FIG. 2.1B

COMBINATION-CONTROL-SURFACE CONCEPT

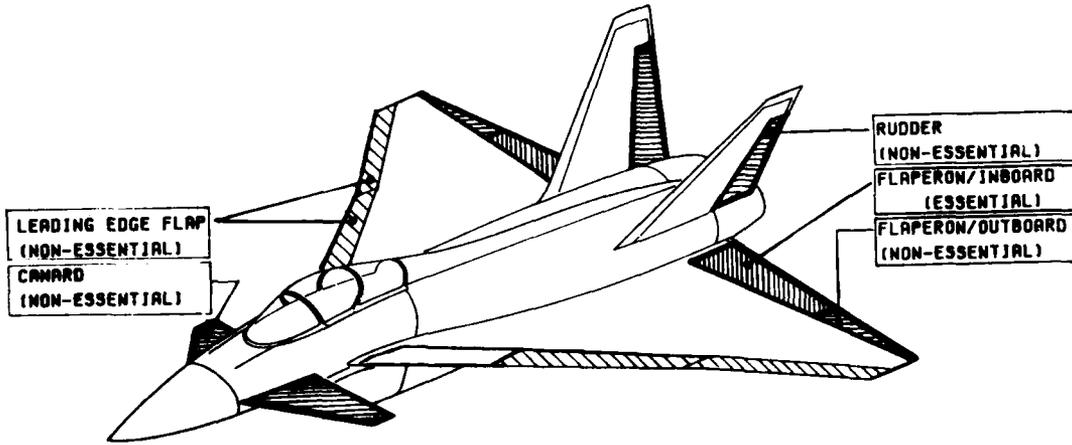


FIG. 2.4A

ACTUATOR CONTROL ARRANGEMENT

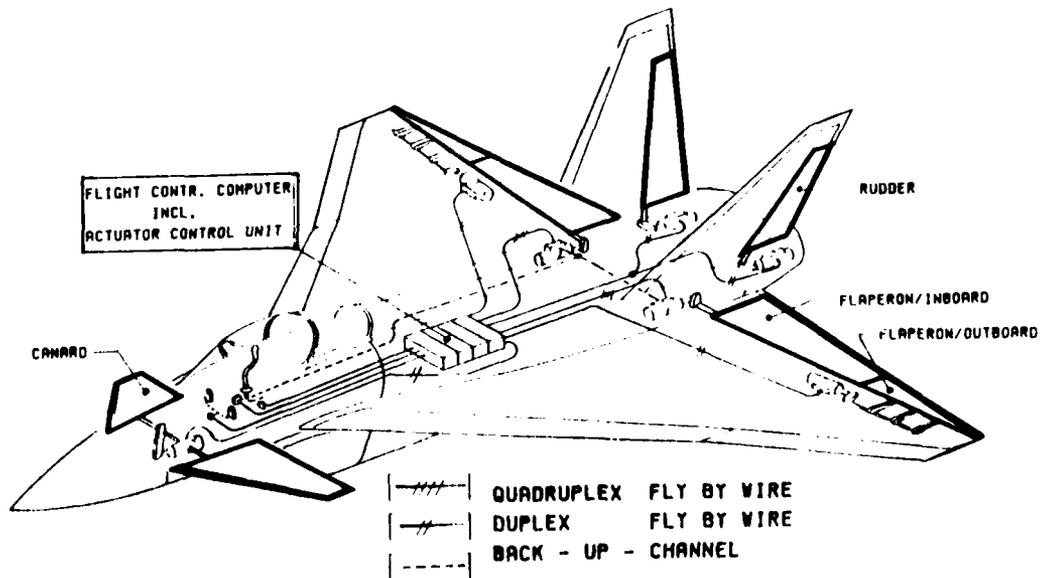


FIG. 2.5A

FLIGHT CONTROL HYDRAULIC ARRANGEMENT

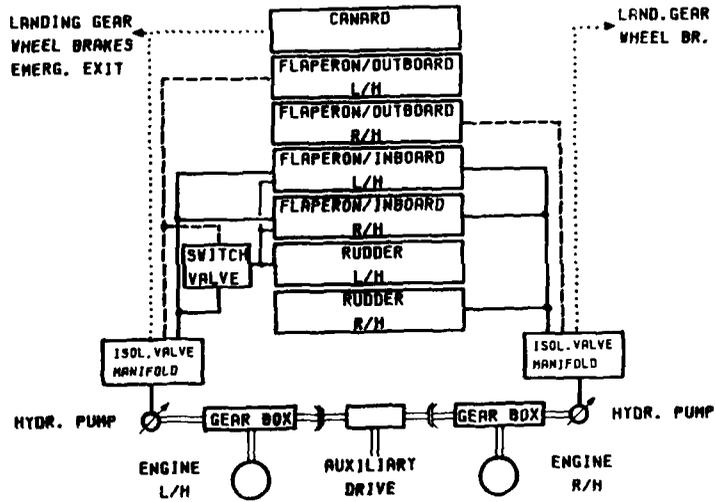


FIG. 2.5B

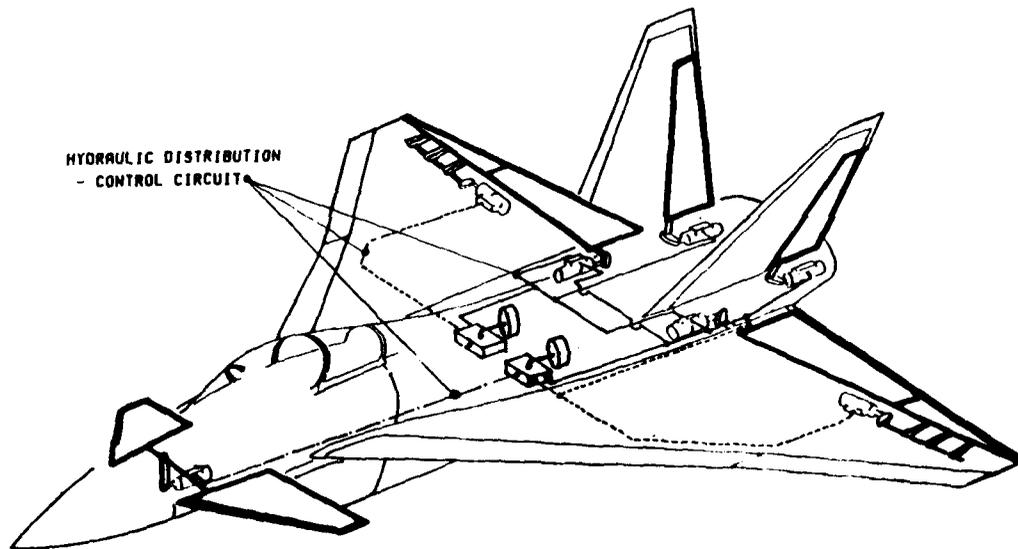


FIG. 2.5C

ACTUATION DESIGN CRITERIA

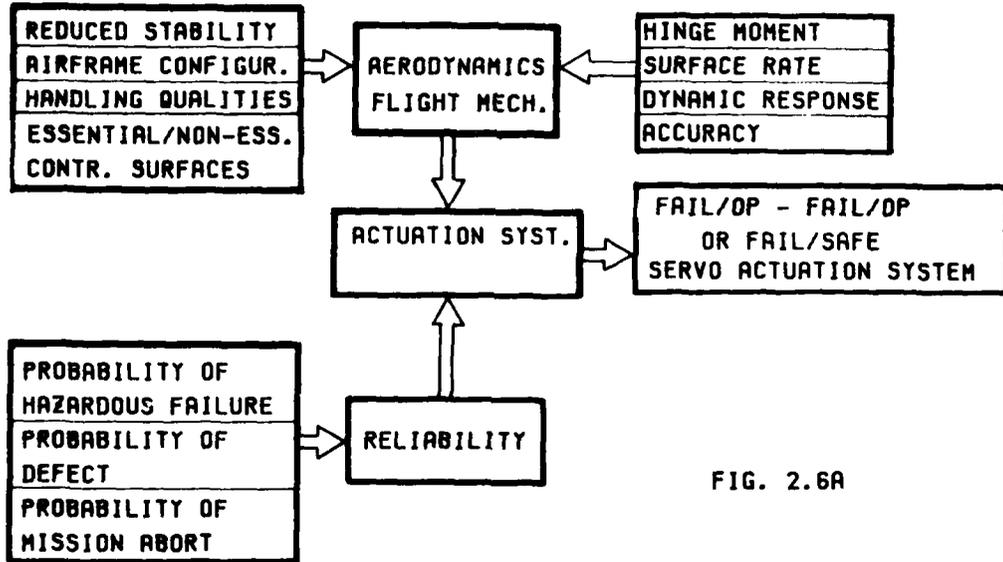


FIG. 2.6A

CONTROL SCHEME FOR A 2 SERVO VALVE ACTUATION SYST.

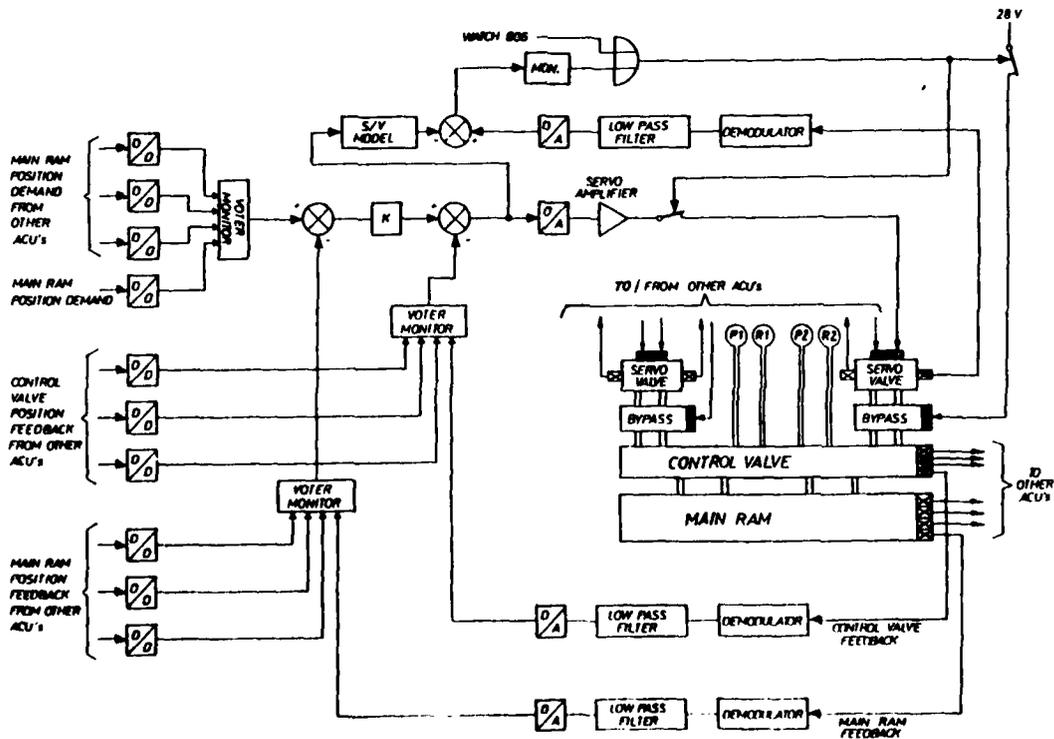


FIG. 2.6B

TWO SERVO VALVE TANDEM ACTUATOR WITH DAMPED FREE FLOATING FAIL SAFE MODE

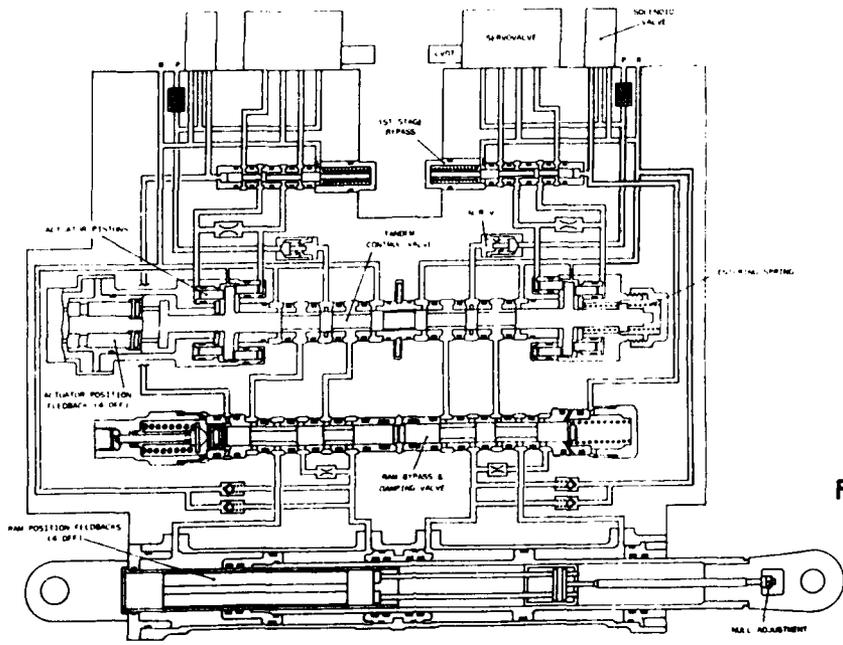


FIG. 2.6C

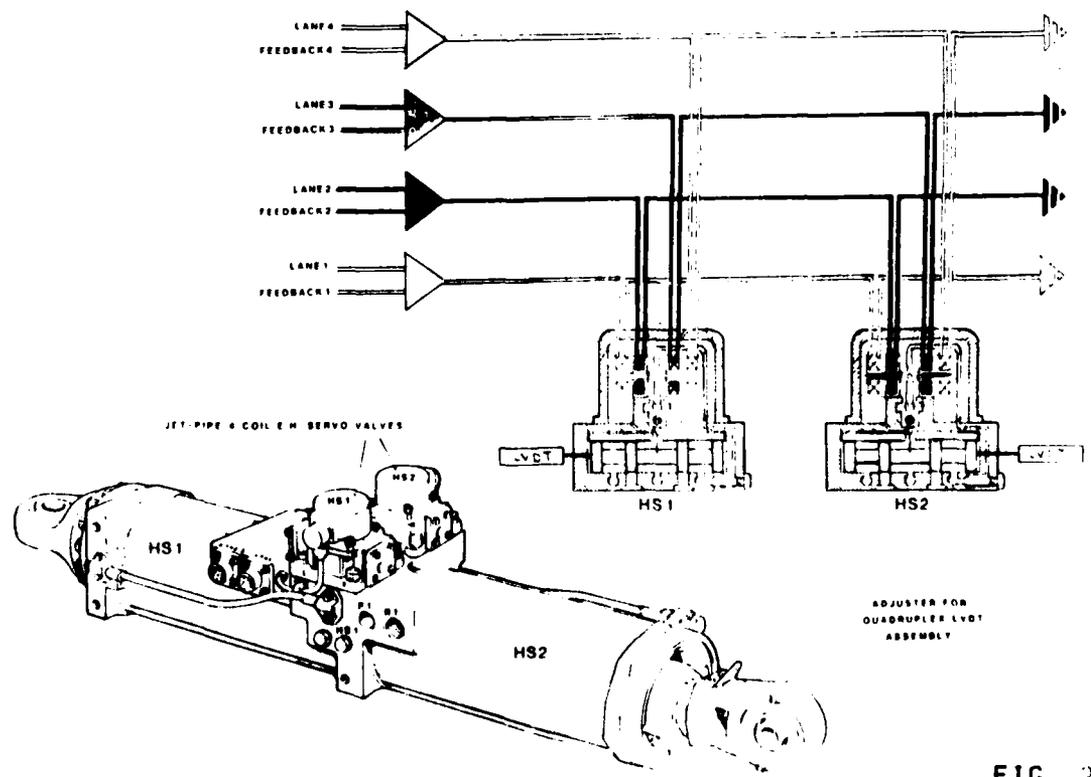


FIG. 2.6D

CONTROL SCHEME FOR A DIRECT DRIVE ACT. SYST.

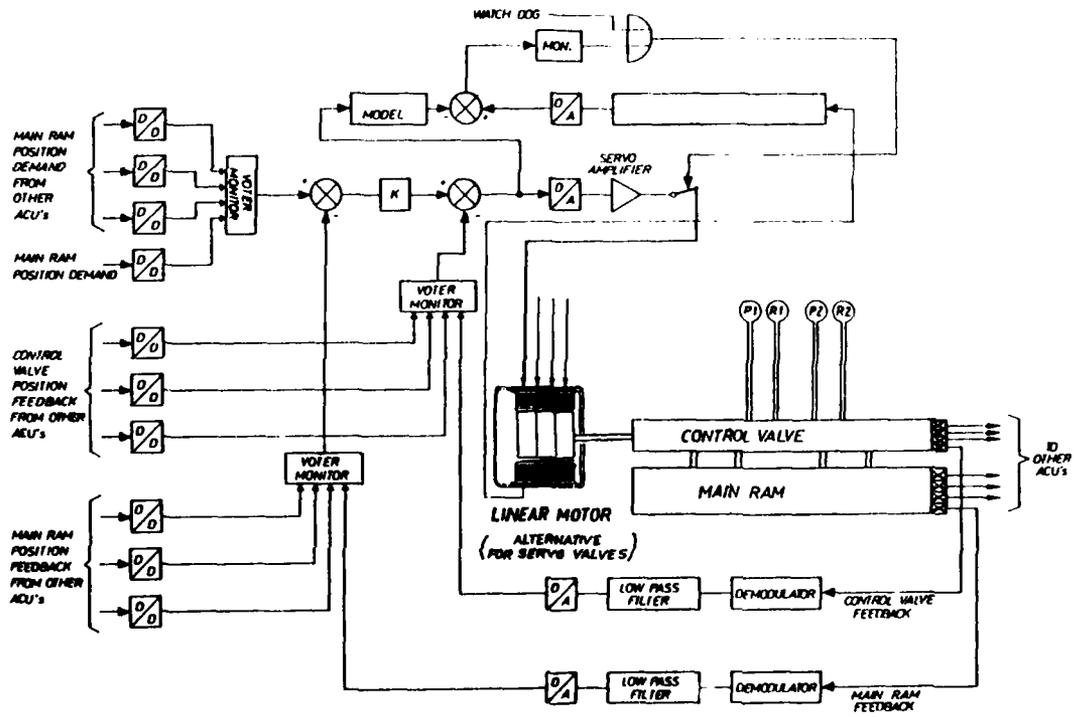


FIG. 2.6E

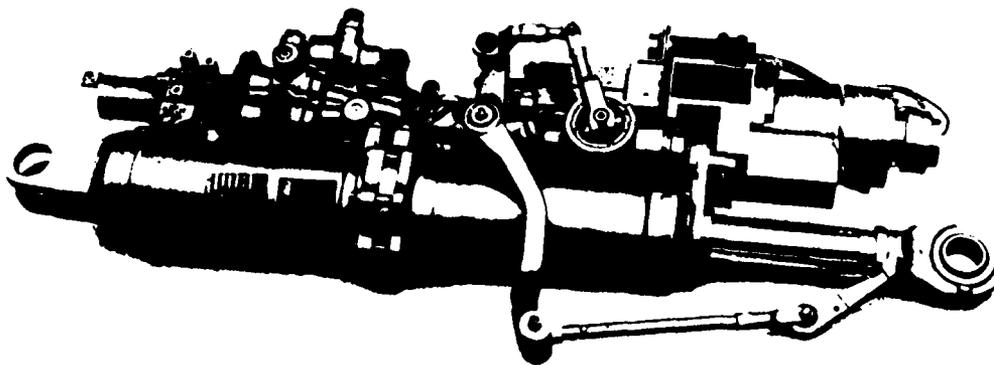


FIG. 2.6F

**ANALYSE COMBINATOIRE PERFORMANCES/COÛTS
D'UN SYSTEME AUTONOME DE NAVIGATION POUR AERONEFS**

par
P. LLORET, G. LAVOPIERRE
S A G E M
6, avenue d'Iéna
75783 PARIS CEDEX 16

(présenté au 39ème Symposium AGARD en Octobre 1984 à CESME - TURQUIE)

RESUME

Cet article propose une méthode simple pour aider les concepteurs de système de navigation autonome pour avion dans le choix d'une solution optimale du point de vue de la performance et du coût. La méthode proposée s'appuie sur le catalogue des performances des systèmes de navigation par inertie et sur des règles originales reliant le prix et la performance ou le volume du système considéré.

L'analyse combinatoire proposée exploite ces données de performances et de coût en comparant le coût d'un système de navigation inertielle considéré comme référence aux coûts de solutions regroupant une solution inertielle moins performante et un ou plusieurs capteurs d'appoint.

Pour illustrer cette méthode, on présente un exemple intéressant : celui de la navigation d'un hélicoptère de combat.

1. INTRODUCTION

Un système de navigation autonome, aussi simple soit-il, utilise la gyroskopie comme référence d'orientation de ses calculs d'estime. Et c'est le plus souvent la qualité de cette fonction gyroskopique qui constitue le facteur clé, à la fois pour les performances obtenues, et pour le coût résultant. Les progrès de la navigation inertielle ont mis à la disposition des "systémiers" un outil qui peut pratiquement apporter une solution complète à tout problème de navigation autonome. Mais les considérations de coût conduisent à différer le plus longtemps possible le recours à cette "panacée", et à essayer de se satisfaire de systèmes gyroskopiques moins performants, mais "aidés" par d'autres références de navigation : essentiellement le cap magnétique et le radar Doppler. Entre la navigation inertielle la plus performante (typiquement, 0,1 à 0,3 Nm/h) et la navigation à l'estime la plus rustique, la qualité de la gyroskopie varie de plus de deux ordres de grandeur, ce qui ne se reflète pas en totalité au niveau des prix, heureusement. L'objet de cette présentation est d'apporter des éléments quantitatifs et logiques sur les rapports qui existent entre les spécifications de performances de l'utilisateur, et le prix du système de navigation autonome optimal qui répond à ces spécifications. Ces considérations se fondent sur les technologies qui constituent "l'état de l'art" en ce milieu des années 1980.

2. DOMAINE COUVERT PAR CETTE PRESENTATION, ET APPROCHE ADOPTÉE

Dans cette présentation, on met principalement l'accent sur les systèmes de navigation à l'estime (dont la navigation inertielle constitue le "haut de gamme"), sans écarter cependant le cas de balisages radioélectriques ou optiques externes à l'aéronef. En particulier, on traite du cas de "navigation relative" que constitue une navigation à l'estime se déroulant entre deux recalages. Les capteurs qui entrent en jeu sont donc gyroskopiques et accélérométriques, d'une part, et capteurs de vitesse et route/cap, de l'autre (radar Doppler ou anémométrie, cap magnétique) pour ce qui est des moyens autonomes, et GPS NAVSTAR, OMEGA, VOR-DME, TAIAN, et recalages optiques, comme aides à une navigation non autonome.

La démarche générale de la présentation est la suivante :

- pour tout système de navigation convenablement spécifié (et on fournit un exemple pour aider l'utilisateur à faire "convenablement" cette spécification), il existe une solution inertielle pure répondant au problème, mais réputée "trop chère", et devant faire l'objet de propositions de remplacement,

- ces solutions de remplacement sont recherchées sous la forme d'une "fonction inertielle" aussi dégradée que possible, mais "aidée" par des capteurs d' "appoint",
- faisant alors le bilan du coût réduit que représente le passage à une "fonction inertielle dégradée", et le coût accru que représente l'adjonction de "capteurs d'appoint", on essaie d'évaluer le bénéfice global,
- cette démarche, supposée s'effectuer de façon itérative entre "client" et "fournisseur" (reprise des spécifications, etc), doit aider à déterminer une solution optimale du point de vue performance/coût.

3. ELEMENT DE BASE : LA PANOPLIE DES SOLUTIONS INERTIELLES

Le tableau est un tableau à double entrée où chaque colonne (repérée de A à E) représente une "classe de performance" de référence inertielle. La première ligne de chaque colonne rappelle le domaine de performance exprimé en erreur de position (milles nautiques par heure à 1 sigma). On voit que seule la partie centrale du "spectre" des performances possibles a été retenue puisque :

- il existe des centrales inertielles trois fois plus précises que celles de la "classe A" dont fait état le tableau,
- à l'autre extrême, il existe des systèmes gyroscopiques beaucoup moins performants que les 30 NM/h dont il est fait état pour la "classe E".

Cependant, le tableau tel qu'il est couvre la grande majorité des cas usuels, et évite une trop grande complexité.

Dans les lignes 2 à 7, on liste les performances qu'il est souhaitable de spécifier pour tout système de navigation autonome, outre la précision de navigation exprimée en milles nautiques par heure. C'est à ce niveau que doit intervenir un "guide du client" pour aider à la spécification. En bref :

La ligne 1 spécifie l'erreur de position absolue (coordonnées terrestres) à long terme souhaitée pour le système de navigation. Ce système pourrait par exemple être à erreur bornée dans le temps, mais très bruitée à court terme. On exprimerait alors dans la ligne 1, à la fois la hauteur du bruit tolérée, et la durée de la mission (supposée durer au moins 1 heure, puisqu'on s'exprime en Nm/h, et que la précision à moyen et court terme est traitée aux lignes suivantes).

La ligne 2 spécifie l'erreur de position absolue (et éventuellement relative) en pourcentage de la distance parcourue. C'est une notion qui couvre à la fois les aspects "long" et "moyen terme" de la navigation, pour un aéronef qui se déplace à une vitesse suffisante (tous aéronefs à voilure fixe, ainsi que la partie "en route" des missions des aéronefs à voilure tournante). Ainsi, pour une centrale inertielle de "classe A", qui accumule en une heure une erreur de 0,3 Nm/h et un déplacement de 300 à 600 Nm, l'erreur relative pourra varier de $0,3/600 = 0,05\%$ à $0,3/300 = 0,1\%$ de la distance parcourue.

Note : on remarquera que par rapport à l'exemple numérique ci-dessus, le tableau (ligne 2, colonne A) fait apparaître une majoration par un facteur 2. Ceci rend compte de l'aspect "moyen terme" de la performance : sur une demi-heure ou un quart d'heure, la vitesse inertielle est plus "bruitée" que son intégrale exprimée en Nm/h

La ligne 3 spécifie l'erreur de position relative à court terme, exprimée en mètres par quart d'heure. C'est une notion qui s'applique principalement aux hélicoptères dans la phase "tactique" de leur mission (déplacements lents).

La ligne 4 spécifie l'erreur de vitesse /sol instantanée. On notera par exemple que pour un système de "classe A", cette erreur n'est pas une simple conversion des Nm/h en m/s. En effet, $0,3 \text{ Nm/h} = 0,15 \text{ m/s}$, alors que c'est une valeur de $0,4 \text{ m/s}$ qui figure dans cette colonne. Ceci rend compte du "bruit de vitesse", de plus en plus "incompressible" à mesure qu'on augmente les performances des centrales inertielles. Noter que cette ligne traite principalement de l'erreur sur le module de la vitesse, l'erreur sur sa direction (l'erreur de route) figurant à la ligne suivante.

La ligne 5 spécifie l'erreur de route instantanée, notion applicable surtout aux aéronefs à voilure fixe, ou à voilure tournante mais dans la phase "en route" de leur mission.

La ligne 6 spécifie l'erreur du cap instantané. Bien qu'il s'agisse d'un paramètre de pilotage, et non de navigation, il est mentionné en tant que "sous produit" possible et spécifiable d'un système de navigation.

La ligne 7 spécifie l'erreur d'attitude instantanée. Là encore, il s'agit d'un paramètre de pilotage qui peut être offert "en prime" par un système de navigation (c'est même l'un des principaux intérêts d'une centrale inertielle de qualité).

Comme il a été dit, spécifier correctement ces 7 paramètres est une tâche pour laquelle le "client" mérite d'être aidé par le fournisseur potentiel.

Le tableau à double entrée étant ainsi défini, on a placé dans chacune de ses cases la valeur typique (ou la plage des valeurs typiques) à laquelle on peut s'attendre pour une centrale inertielle pure de la classe considérée (A à E). En d'autres termes, les chiffres qui figurent dans une colonne donnée sont cohérents entre eux pour une centrale inertielle de la classe considérée. Ce qui ne signifie pas que l'utilisateur potentiel ne peut fournir de spécifications que dans une seule colonne. Au contraire, certaines valeurs spécifiées vont apparaître comme étant plus contraignantes que d'autres, et c'est de cette constatation que partira la tentative d'optimisation (qui consiste, rappelons-le, à remplacer de la "gyroscopie chère" par de la gyroscopie moins chère mais aidée, si possible, et voir en fin de compte si on gagne au change).

4. ANALYSE DES PERFORMANCES ET DES SOLUTIONS TECHNIQUES ASSOCIEES

4.1 PERFORMANCES SPECIFIABLES, ET POSITIONS DE REPLI (figures 1 et 2)

Une première lecture du tableau peut se faire comme suit :

- dans les colonnes de gauche (en gros, A, C), on trouve des performances relativement élevées qui correspondront vraisemblablement à l'objectif poursuivi par le client. Ce sont les "performances spécifiables".
- dans les colonnes de droite (en gros, D, E), on trouve des performances relativement basses que le client demandera rarement en tant que telles, mais dont "il pourra se contenter" (position de repli) si on lui démontre qu'avec l'aide de capteurs d'appoint, on peut passer d'une centrale inertielle de classe élevée à une centrale de classe plus faible (et idéalement, le plus faible possible, pour des raisons de coût).

Plus précisément la répartition suivante est très probable :

Performances spécifiables :

Colonne A et colonne B entières
Colonne C à l'exception de C3 éventuellement
Cases D6 et D7

Performances "de repli":

Colonne E entière
Cases D1, D2, D3, D4.

4.2 PERFORMANCES SPECIFIABLES NECESSITANT UNE CENTRALE INERTIELLE (figure 3)

Si, dans le tableau, le client spécifie une ou plusieurs des "cases" que nous allons lister ci-après, il spécifie du même coup un système inertiel de la classe de performance correspondant à la colonne cochée, sans qu'il y ait d'alternative possible. Les raisons n'en sont cependant pas toujours les mêmes :

Cases 4A, 5A, 5B, 5C, 6A, 6B, 7A, 7B, 7C

Dans chacun de ces cas, la centrale inertielle est le seul capteur à pouvoir fournir le paramètre avec la précision indiquée. Plus précisément :

4A : tout radar Doppler est trop bruité pour pouvoir fournir une vitesse instantanée aussi lisse ; la centrale inertielle est indispensable, ne serait-ce que pour "liaser" le Doppler.

5A, 5B, 5C : la centrale inertielle est la seule référence de route à bord ; à défaut, on peut utiliser une référence de cap, mais en faisant l'hypothèse d'un angle de dérive nul, ce qui peut occasionner une erreur conséquente (et probablement même supérieure à la valeur qui se trouve dans la case 5B, qu'on pourrait alors rattacher aux précédentes).

6A, 6B : pour le cap (géographique), la seule alternative à la centrale inertielle est un cap gyromagnétique, dont la précision (en termes de cap géographique) n'atteindra pas mieux que le degré.

7A, 7B, 7C : pour une précision d'attitude meilleure que 0,5°, il n'y a pas d'alternative aux centrales inertielles.

4.3 SOLUTIONS DE REPLI SORTANT DU CADRE D'UNE NAVIGATION AUTONOME (figure 4)

Parmi les performances qu'on peut spécifier à l'aide du tableau, il en est qui peuvent être atteintes autrement qu'avec une navigation autonome : ce sera essentiellement avec l'aide d'une radiolocalisation ou d'une localisation par voie optique. Il en est ainsi pour les lignes 1, 2 et 3 du tableau. In effet :

Ligne 1 : si on se contente de spécifier une précision de position, un moyen de radio-localisation, ou de recalage optique, pourra toujours faire l'affaire. Certes, des aides du type VOR-DME ou OMEGA auront des difficultés à garantir une précision équivalente à celle des centrales inertielles de classe A, B (ou même C) sur des durées de mission de une à quelques heures. Mais on pourra alors songer au GPS NAVSTAR ou aux recalages optiques. Donc, pour une spécification située sur cette ligne, on peut toujours trouver une solution autre qu'autonome, mais qui sort par conséquent du cadre de la présente étude.

Ligne 3 : même situation, globalement, que pour la ligne 1. Noter que la précision de 200 mètres sur 15 minutes est déjà difficile à obtenir avec les meilleures aides radioélectriques à la navigation (GPS NAVSTAR en accès "tous utilisateurs").

Ligne 2 : pourrait être assimilée aux deux lignes précédentes, à ceci près que la spécification en % de la distance parcourue favorise les solutions inertielles, car elle est draconienne pour les aides radioélectriques bruitées à court terme. Si on veut réellement spécifier des précisions en % de la distance parcourue, il est beaucoup de cas où la seule solution est inertielle (ne serait-ce que parce qu'un angle de route précis est nécessaire).

4.4 PERFORMANCES POUVANT ETRE SATISFAITES SANS CENTRALE INERTIELLE (figure 5)

Ce sont essentiellement celles des colonnes D et E du tableau. En d'autres termes, ce que peuvent procurer les centrales inertielles de bas de gamme peut aussi être obtenu par d'autres moyens. Par exemple :

Cap et attitude instantanés (lignes 6 et 7) : au niveau de performances de 2° (case D6) à 7° (case E6) en cap, ou de 1° (case D7) à 3° (case E7) en attitude, les gyroscopes directionnels et de verticale répondent à la question.

Route instantanée (ligne 5) : au niveau de performances de 7° à 20° (case E5) en route instantanée, assimiler route et cap (obtenu par référence magnétique) répond à la question. Ce qui est déjà beaucoup plus sujet à caution pour une précision de 2° à 7° (case D5).

Vitesse/sol instantanée (ligne 4) : pratiquement toutes les performances de la ligne 4 peuvent être apportées par un radar Doppler, à l'exception des 0,4 m/s de la case A4, et ceci plus pour des raisons de bruit à court terme du radar Doppler, que d'erreur moyenne de ce dernier (nécessité d'un lissage inertielle). En outre, cette ligne ne concerne que le module de la vitesse, ce qui est rarement utile sans information de route (ligne 5) où on retrouve le besoin d'une centrale inertielle.

Position relative à court terme (ligne 3) : à partir de la performance 1800m / 15 minutes (colonnes C,D,E), un système de type radar Doppler + cap gyromagnétique est suffisant.

Précision de position à moyen terme (ligne 2) : seules les performances des colonnes D et E (donc, plus de 3 % de la distance parcourue) peuvent être atteintes sans risque par des moyens non inertielles mais autonomes (Doppler plus cap gyromagnétique). La performance C2 (1 à 2 % de la distance parcourue) n'est obtenue que marginalement, car elle est très exigeante pour le cap gyromagnétique utilisé en capteur de cap géographique (adjonction de la déclinaison magnétique, perturbations magnétiques diverses etc).

Précision de position à long terme (ligne 1) : si on reste dans le domaine de la navigation autonome, seule la performance E1 (30 Nm/h) et marginalement la D1 (10 Nm/h) peuvent être obtenues par des moyens non inertielles.

4.5 RAPPEL DES CAS MARGINAUX

Ce rappel est intéressant car il situe la ligne de partage des performances entre solutions inertielles et non-inertielles, et marque donc une "ligne de front" dans la bataille pour l'optimisation du rapport performances / prix. Ainsi, il apparaît que le besoin d'une centrale inertielle commence à se faire sentir quand on en arrive à demander des performances égales ou supérieures à :

- 10 Nm/h en précision de la position absolue à long terme,
- 1 à 2 % d'erreur de position relative à la distance parcourue,
- 600 m d'erreur de position relative pour 15 minutes de mission dans une zone tactique de dimensions limitées,

et, en ce qui concerne les paramètres de pilotage :

- 2° à 7° pour la précision de la route instantanée,
- 1° sur la précision du cap instantané,
- 0,5° sur la précision de l'attitude instantanée.

4.6 PERFORMANCES SPECIFIABLES POUVANT ETRE ATTEINTES, SOIT EN "INERTIE PURE", SOIT EN "INERTIE DEGRADEE AIDEE" (figure 6)

On liste ci-après les "cas de base" (inertie pure) et les solutions de repli possibles, en se référant aux cases du tableau. Ainsi, par exemple, la case A1 (0,3 Nm/h) affiche une performance qui peut être atteinte en inertie pure, mais également avec un GPS filtré par une centrale inertielle de classe E, ou par des recalages optiques avec interpolation assurée par centrale inertielle de classe C ou D. On notera ceci :

A1 -----> * GPS + E
ou * Optique + C (ou D) et de même:

B1 -----> * OMEGA ou VOR/DME ou TACAN + E
ou * Optique + C (ou D)

C1 -----> * OMEGA ou VOR/DME ou TACAN + "F"
ou * Optique + D

D1 : rappel : * marginalement réalisable par Doppler + cap gyromagnétique
* peu demandé en tant que "spécification positive"

De même pour la ligne 2 :

A2 -----> * GPS + C (ou D)
ou * Optique + B

B2 -----> * Doppler + C
ou * Doppler + optique + D (ou E)

C2 -----> * Doppler + D
ou * Doppler + optique + E

ou (rappel) : marginalement réalisable avec Doppler + cap gyromagnétique

De même pour la ligne 3 :

A3 -----> * GPS + E
ou * Optique + C (ou D)

B3 -----> * GPS + E
ou * Optique + C (ou D)

ou (rappel) : marginalement réalisable avec Doppler + cap gyromagnétique.

4.7 RECAPITULATION DE L'ANALYSE COMBINATOIRE DES PERFORMANCES REALISEE CI-DESSUS

1. Dans le tableau, on a recensé les paramètres de performances spécifiés (lignes) et leurs ordres de grandeur numériques pour le cas où on se propose de les obtenir en inertie pure. Le tableau permet donc, en pointant les cases correspondant aux performances demandées, de savoir quelle classe de centrale (notée de A à E) serait nécessaire pour obtenir en inertie pure lesdites performances.
2. Dans le § 4.2 (ci-dessus), on a recensé celles de ces performances qui, de toute façon, ne peuvent être obtenues qu'en inertie pure (pas de "solution de repli").
3. Dans le § 4.3 (ci-dessus), on a recensé celles des performances qui peuvent être atteintes en recourant à des aides à la navigation externes (non autonomie). Il s'agit donc d'un sous-ensemble du groupe de performances pour lesquelles on peut envisager une "solution de repli" autre qu'inertielle.
4. Dans le § 4.4 (ci-dessus), on a recensé celles des performances qui peuvent être atteintes sans recours à une centrale inertielle, et dans le paragraphe suivant (5.4.5), on a passé en revue les "performances marginales" à partir desquelles l'introduction d'une centrale inertielle commence à se faire sentir.
5. Enfin, dans le § 4.6 (ci-dessus), on a recensé celles des performances qui peuvent être atteintes, soit en inertie pure, soit en "inertie dégradée aidée", ce qui donne alors matière à compromis et optimisation.

A cette combinatoire des performances, il faut maintenant ajouter la combinatoire des coûts correspondants, objet des paragraphes suivants.

5. COUT DES CENTRALES INERTIELLES EN FONCTION DE LEURS PERFORMANCES

Les considérations qui suivent s'appuient sur une certaine connaissance des prix de marché qui, à long terme, reflètent les coûts "physiques". On n'a donc pas essayé de procéder de manière analytique, en examinant le contenu des systèmes et les technologies en cause. Mais ces considérations technologiques sont englobées et sous-jacentes dans la présentation qui est faite des classes de performances des centrales inertielles. Par exemple :

- pour les centrales de classe A, on sait que les plates-formes à gyroscopes accordés de haute précision, ou avec gyroscopes à suspension électrique, répondent à la question, pour un certain prix de marché. C'est ce prix de marché qui a été évalué, sans qu'on cherche ici à savoir laquelle des deux solutions l'atteint... avec le meilleur profit pour le fournisseur.
- pour les centrales de classe B, on sait que les plates-formes à gyroscopes accordés, ou les centrales à gyrolasers, répondent à la question. Là encore, on connaît un prix de marché, qui couvre les deux cas de technologie pour l'usage que nous avons à en faire.
- les centrales du groupe C peuvent être à plate-forme avec gyroscopes accordés, ou "strap-down" à gyroscopes accordés, ou gyrolaser de périmètre réduit etc. Le prix de marché correspondant à cette classe de centrales a été par définition posé égal à 1, dans la mesure où il représente le "cas central" du tableau et de la gamme de performances qu'il représente. Donc :

L'unité de coût utilisée pour le suite de la présentation est le coût d'une centrale inertielle de "classe C" selon la définition du tableau.

- les centrales des groupes D et E sont en fait plutôt des systèmes gyroscopiques de relativement basses performances, auxquels on ne demande pas vraiment un fonctionnement autonome, mais plutôt une fonction de lissage ou d'interpolation entre recalages. Le tableau mentionne néanmoins leurs "performances fictives" en inertie pure.

Ayant ainsi fixé l'étalon de coût comme étant celui d'une centrale de classe C, le premier pas consiste à trouver une formule donnant le coût relatif des centrales de classes A, B (coût > 1) et D, E (coût < 1). En posant que :

- $(Nm/h_0) = 3$ (milles nautiques par heure) = précision de la centrale de classe C
- (Nm/h) = précision en Nm/h de la centrale dont on veut connaître le coût
- $C_0 = 1$ = coût de la centrale de classe C
- C = coût de la centrale définie par son paramètre de précision (Nm/h)

Nous proposons la formule suivante :

$$C/C_0 (= C/1 = C) = [(Nm/h_0)/(Nm/h)]^x$$

avec x compris entre 0,30 et 0,35

Le tableau qui suit donne les valeurs numériques pour deux valeurs extrêmes et une valeur médiane de x , à titre de "jeu d'essai" :

| CLASSE DE PERFORMANCE | A | B | C | D | E |
|-----------------------|------|------|---|------|------|
| Nm/h | 0,3 | 1 | 3 | 10 | 30 |
| Coût avec $x = 0,30$ | 1,99 | 1,39 | 1 | 0,7 | 0,5 |
| Coût avec $x = 0,33$ | 2,13 | 1,44 | 1 | 0,67 | 0,47 |
| Coût avec $x = 0,36$ | 2,29 | 1,48 | 1 | 0,65 | 0,44 |

L'idée qui se dégage de ce tableau, c'est que lorsqu'on augmente les performances d'un demi ordre de grandeur (facteur de l'ordre de 3), les coûts sont augmentés de 40 % environ. Pour gagner un ordre de grandeur (facteur 10), il faut plus que doubler le coût. Certes, cette loi du type $(\log a) = k \cdot (\log b)$ n'a rien de physique, mais elle rend approximativement compte d'une relation coût/performance sur l'étendue de deux décades (facteur 100 sur les performances). Et elle rappelle par ailleurs les lois de Wright qui s'appliquent aux décroissances de coût de production en fonction du rang (le coût est multiplié par $x < 1$ chaque fois que le rang double) et qui, bien que n'ayant rien de physique, rendent néanmoins d'estimables services dans les évaluations de coûts.

5.1 UTILISATION DIRECTE DE LA FORMULE COÛT/PERFORMANCES POUR CENTRALES INERTIELLES

En regardant le tableau du § 5 (ci-dessus), le spécificateur voit directement ce qu'il en coûte de spécifier une performance supérieure par exemple d'un demi-ordre de grandeur à ce qui est strictement nécessaire. Utilisant le tableau, il est aussi en mesure de noter quel est le paramètre spécifié qui est "dimensionnant" en matière de coût (en d'autres termes, quel est le paramètre "meneur"), et pourra s'attacher à le "faire rentrer dans le rang" pour pouvoir choisir une centrale de classe inférieure.

Exemple : si au lieu d'une erreur de route instantanée de 0,2° à 0,7° (paramètre B5), le spécificateur peut se contenter d'une précision comprise entre 0,7° et 2° (paramètre E5), le coût passera de 1,44 à 1, ce qui n'est pas du tout négligeable.

5.2 UTILISATION COMBINEE DE LA FORMULE COUT/PERFORMANCES POUR CENTRALES INERTIELLES, ET DES ALTERNATIVES "INERTIE PURE/INERTIE AIDEE" DU § 4.

Prenons l'exemple de la spécification A1 (0,3 Nm/h en position à long terme) qui peut être satisfaite, soit par une centrale inertielle de classe A, soit par le GPS associé à une "centrale inertielle" de classe E. Le bilan de coût peut se faire de la façon suivante :

Coût d'une centrale inertielle de classe A : 2,13

Coût d'une centrale inertielle de classe E : 0,47

Crédit ("Economie réalisée") : 1,66

Débit : coût global d'un équipement GPS (antenne, récepteur etc).

Evidemment, il s'agit là d'un exemple simpliste réduit à un seul paramètre. En réalité, il est probable qu'on voudra spécifier, non seulement la position, mais aussi la vitesse/sol (module) et la route. Ce qui a été développé précédemment permet de procéder à ce genre d'itérations.

UN CAS D'ECOLE INTERESSANT : LA NAVIGATION DE L'HELICOPTERE DE COMBAT

Le "portrait-type" d'une spécification pour hélicoptère de combat serait, dans le langage codé que permet l'utilisation du tableau :

C2 (navigation en route = 1% à 2% de la distance parcourue)

B3 (navigation tactique 600m/15min) ou même deux fois mieux si possible (A3?)

B4 (vitesse/sol instantanée 1 m/s pour raisons autres que la navigation)

...et autres paramètres moins déterminants pour notre exemple. A partir de cette spécification, on peut procéder comme suit :

1. S'il fallait procéder en inertie pure, il faudrait une centrale de classe B, dont le coût est de l'ordre de 1,44. Ceci à cause de deux paramètres (lignes 3 et 4 du tableau).
2. Cependant, on sait que le paramètre N° 3 (position relative à court terme) peut être obtenu avec une centrale de classe E aidée par GPS, ou de classe C (ou D) aidée par recalages optiques de position. Quant au paramètre N° 4 (module de la vitesse/sol), le radar Doppler suffit. Quant à la spécification "de second rang" C2 (navigation à 1-2% de la distance parcourue), elle peut aussi être satisfaite par une centrale de classe D aidée par Doppler, ou bien E aidée par Doppler et recalages optiques. De la combinatoire de tout ceci, il ressort que :
3. On peut introduire le Doppler pour satisfaire à la spécification B4, qui n'implique plus alors pour elle-même la présence d'une centrale inertielle.
4. Mais cette centrale inertielle reste tout de même nécessaire pour satisfaire à la spécification B3. Dans ce cas, deux possibilités :
 - il suffira qu'elle soit de classe E si on a par ailleurs le GPS,
 - il faudra qu'elle soit de classe C (ou D, à la rigueur) si on veut procéder par recalages optiques, en lieu et place de GPS.
5. Par ailleurs, le Doppler, supposé introduit pour satisfaire à B4, permet de passer d'une centrale de classe C à classe D (ou même E) pour satisfaire à la spécification C2.

Ainsi se trouvent confrontées les possibilités suivantes :

- centrale classe B,
- Doppler + recalages optiques + centrale classe D (ou E),
- GPS (fournissant aussi vitesse) + centrale classe E (revenir là-dessus "en amont").

Et le problème sera de savoir si le passage d'une centrale inertielle de classe B à une centrale de classe D "paie" le coût d'une Doppler et des recalages optiques de navigation, ou si le passage d'une centrale de classe B à une gyroscopie de classe E "paie encore mieux" le coût d'une GPS et des recalages optiques.

En première approximation, on peut dire ce qui suit :

A. Comparaison centrale classe B/centrale classe D + Doppler + recalage opt.

Différentiel de coût centrale classe B - centrale classe D :

$$1,44 - 0,67 = 0,77$$

Pour que la solution à radar Doppler + recalages de position soit avantageuse, il faut que le coût de ce radar et du dispositif de recalage optique associé soit inférieur à 0,77 (soit 0,77 fois le coût d'une centrale moyenne de la classe C, c'est-à-dire 3 Nm/h). En principe, c'est le cas si on ne considère que le coût du Doppler sans les frais associés (nécessité d'une ouverture ventrale etc), et si on compte pour rien le coût des recalages optiques ("il suffit de" passer sur des repères terrestres convenus à l'avance). C'est la raison pour laquelle cette solution est habituellement retenue. Elle présente peut-être des "coûts opérationnels" moins facilement quantifiables (aléas accrus sur la position, résultant des aléas de recalages, charge de travail accrue etc), mais ceci sort du cadre de notre étude.

B. Comparaison centrale classe B/centrale classe E + GPS

Différentiel de coût centrale classe B - centrale classe E :

$$1,44 - 0,47 = 0,97$$

Pour que la solution à GPS soit avantageuse, il faut que le coût du GPS et de son installation (implantation et asservissement d'une antenne directive en un lieu favorable de l'appareil etc) soit inférieur à 0,97, soit pratiquement celui d'une centrale inertielle de classe C (3 Nm/h). Outre ce coût financier, le spécificateur considérera aussi le coût opérationnel d'avoir un système dont l'autonomie est réduite à presque zéro (tout au plus quelques minutes en cas de perte du GPS). Mais là encore, ceci n'est pas quantifiable et ne peut donc être pris en compte dans cette étude.

Nota : Une solution avec Doppler + GPS + "centrale inertielle" classe E est en principe également envisageable. Il faudrait alors que le coût combiné du Doppler et du GPS soit inférieur à 0,97, ce qui semble très improbable.

5.3 AUTRES FACTEURS FONDAMENTAUX POUR UNE ETUDE DE COÛTS : LE VOLUME DES EQUIPEMENTS, ET LA QUANTITE QUI EN EST PRODUITE

Dans tout ce qu'on a considéré jusqu'ici, les centrales inertielles étaient caractérisées par un seul paramètre : leurs performances. Ce qui faisait implicitement l'hypothèse : "toutes choses égales par ailleurs". Parmi les choses qui peuvent ne pas être égales par ailleurs, il y en a au moins deux dont l'incidence sur les coûts est primordiale, et en tout cas d'ordre de grandeur comparable à ce dont on vient de traiter jusqu'ici : il s'agit du volume alloué pour l'équipement, et de la quantité d'équipements à produire.

En ce qui concerne les quantités, chacun est maintenant familiarisé avec les formules de Wright, du type :

$$C/C_0 = (Q_0/Q)^y$$

avec :

- C_0 = coût pour la quantité Q_0
- C = coût pour la quantité Q
- y (exposant) = paramètre de décroissance des coûts en fonction de la quantité, exprimant par exemple que les coûts sont multipliés par 0,9 quand les quantités doublent.

Dans le cas des centrales inertielles, une telle formule peut être considérée comme applicable dans les limites et avec les ordres de grandeur suivants :

- $Q_0 = 500$
- Q compris entre 50 et 5000
- y compris entre 0,07 (les coûts baissent de 5 % quand les quantités doublent) et 0,24 (les coûts baissent de 15 % quand les quantités doublent), avec une valeur médiane autour de 0,15.

En ce qui concerne les volumes alloués pour les centrales inertielles, l'incidence sur les coûts est très importante. On peut admettre une formule du type :

$$C/C_0 = (V_0/V)^z$$

applicable dans les limites et avec les ordres de grandeur suivants :

- $V_0 = 17$ litres (pour l'unité inertielle)
- V compris entre 0,7 et 1,5 V_0
- z compris entre 1,2 et 1,6, avec une valeur probable autour de 1,4 (le coût est multiplié par 1,6 lorsque le volume est multiplié par 0,7).

Encore une fois, ces considérations recouvrent une variété de technologies diverses et leurs prix de marché. Dans certains cas, telle technologie aura "du mal à suivre" (par exemple le gyrolaser pour les faibles volumes et les hautes performances).

mais on trouvera toujours un "prix de marché" pour le volume et la performance spécifiés, et c'est ce dont la formule se propose de rendre compte.

Globalement, pour tenir compte des principaux paramètres de coût, on pourra donc adopter pour les centrales inertielles une formule du type suivant :

$$C/C_0 = [(Nm/h)_0 / (Nm/h)]^x \cdot (Q_0/Q)^y \cdot (V_0/V)^z$$

Dans les limites et avec les ordres de grandeur spécifiés plus haut.

TABLEAU : CLASSIFICATION DES CENTRALES INERTIELLES EN CLASSES DE PERFORMANCES

| CLASSE DE PERFORMANCE | | A | B | C | D | E |
|-----------------------|---|----------|---------|-------|------|-------|
| Nm/h | 1 | 0,3 | 1 | 3 | 10 | 30 |
| % x d | 2 | 0,1-0,2 | 0,3-0,6 | 1-2 | 3-6 | 10-20 |
| m/15 min | 3 | 200 | 600 | 1800 | 6000 | 18000 |
| m/s | 4 | 0,4 | 1,3 | 4 | 13 | 40 |
| degrés (route) | 5 | 0,07-0,2 | 0,2-0,7 | 0,7-2 | 2-7 | 7-20 |
| degrés (cap) | 6 | 0,1 | 0,3 | 1 | 2 | 7 |
| degr. (attitude) | 7 | 0,05 | 0,1 | 0,3 | 1 | 3 |

FIGURE 1 - PERFORMANCES SPECIFIABLES

| CLASSE DE PERFORMANCE | | A | B | C | D | E |
|-----------------------|---|----------|---------|-------|------|-------|
| Nm/h | 1 | 0,3 | 1 | 3 | 10 | 30 |
| % x d | 2 | 0,1-0,2 | 0,3-0,6 | 1-2 | 3-6 | 10-20 |
| m/15 min | 3 | 200 | 600 | 1800 | 6000 | 18000 |
| m/s | 4 | 0,4 | 1,3 | 4 | 13 | 40 |
| degrés (route) | 5 | 0,07-0,2 | 0,2-0,7 | 0,7-2 | 2-7 | 7-20 |
| degrés (cap) | 6 | 0,1 | 0,3 | 1 | 2 | 7 |
| degr. (attitude) | 7 | 0,05 | 0,1 | 0,3 | 1 | 3 |

FIGURE 2 - PERFORMANCES DE REPLI

| CLASSE DE PERFORMANCE | | A | B | C | D | E |
|-----------------------|---|----------|---------|-------|------|-------|
| Nm/h | 1 | 0,3 | 1 | 3 | 10 | 30 |
| % x d | 2 | 0,1-0,2 | 0,3-0,6 | 1-2 | 3-6 | 10-20 |
| m/15 min | 3 | 200 | 600 | 1800 | 6000 | 18000 |
| m/s | 4 | 0,4 | 1,3 | 4 | 13 | 40 |
| degrés (route) | 5 | 0,07-0,2 | 0,2-0,7 | 0,7-2 | 2-7 | 7-20 |
| degrés (cap) | 6 | 0,1 | 0,3 | 1 | 2 | 7 |
| degr. (attitude) | 7 | 0,05 | 0,1 | 0,3 | 1 | 3 |

FIGURE 3 - PERFORMANCES SPECIFIABLES NECESSITANT UNE CENTRALE INERTIELLE

| CLASSE DE PERFORMANCE | | A | B | C | D | E |
|-----------------------|---|----------|---------|-------|------|-------|
| Nm/h | 1 | 0,3 | 1 | 3 | 10 | 30 |
| % x d | 2 | 0,1-0,2 | 0,3-0,6 | 1-2 | 3-6 | 10-20 |
| m/15 min | 3 | 200 | 600 | 1800 | 6000 | 18000 |
| m/s | 4 | 0,4 | 1,3 | 4 | 13 | 40 |
| degrés (route) | 5 | 0,07-0,2 | 0,2-0,7 | 0,7-2 | 2-7 | 7-20 |
| degrés (cap) | 6 | 0,1 | 0,3 | 1 | 2 | 7 |
| degr. (attitude) | 7 | 0,05 | 0,1 | 0,3 | 1 | 3 |

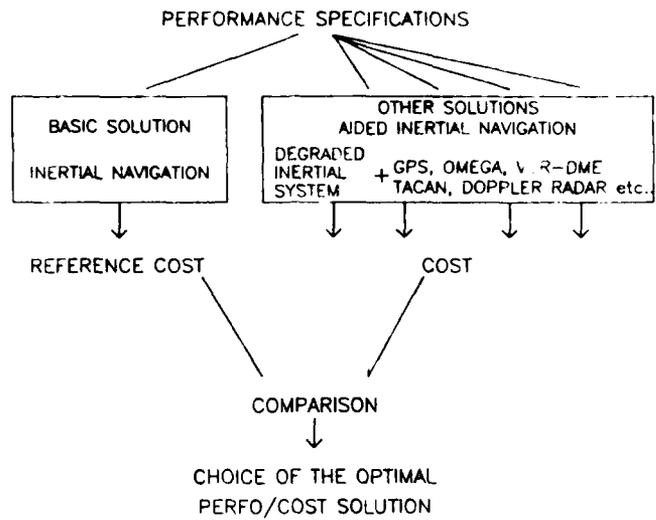
FIGURE 4 - SOLUTION DE REPLI SORTANT DU CADRE D'UNE NAVIGATION AUTONOME

| CLASSE DE PERFORMANCE | | A | B | C | D | E |
|-----------------------|---|----------|---------|-------|------|-------|
| Nm/h | 1 | 0,3 | 1 | 3 | 10 | 30 |
| % x d | 2 | 0,1-0,2 | 0,3-0,6 | 1-2 | 3-6 | 10-20 |
| m/15 min | 3 | 200 | 600 | 1800 | 6000 | 18000 |
| m/s | 4 | 0,4 | 1,3 | 4 | 13 | 40 |
| degrés (route) | 5 | 0,07-0,2 | 0,2-0,7 | 0,7-2 | 2-7 | 7-20 |
| degrés (cap) | 6 | 0,1 | 0,3 | 1 | 2 | 7 |
| degr. (attitude) | 7 | 0,05 | 0,1 | 0,3 | 1 | 3 |

FIGURE 5 - PERFORMANCES POUVANT ETRE SATISFAITES SANS CENTRALE INERTIELLE

| CLASSE DE PERFORMANCE | | A | B | C | D | E |
|-----------------------|---|----------|---------|-------|------|-------|
| Nm/h | 1 | 0,3 | 1 | 3 | 10 | 30 |
| % x d | 2 | 0,1-0,2 | 0,3-0,6 | 1-2 | 3-6 | 10-20 |
| m/15 min | 3 | 200 | 600 | 1800 | 6000 | 18000 |
| m/s | 4 | 0,4 | 1,3 | 4 | 13 | 40 |
| degrés (route) | 5 | 0,07-0,2 | 0,2-0,7 | 0,7-2 | 2-7 | 7-20 |
| degrés (cap) | 6 | 0,1 | 0,3 | 1 | 2 | 7 |
| degr. (attitude) | 7 | 0,05 | 0,1 | 0,3 | 1 | 3 |

FIGURE 6 - PERFORMANCES POUVANT ETRE ATTEINTES, SOIT EN INERTIE PURE, SOIT EN INERTIE DEGRADEE AIDEE



INERTIAL SYSTEM CLASSIFICATION ACCORDING TO PERFORMANCE

| PERFORMANCE CLASS | | A | B | C | D | E | |
|--|---|----------|---------|-------|------|-------|---------|
| Long-term position accuracy | 1 | 0.3 | 1 | 3 | 10 | 30 | Nm/h |
| Medium-term position accuracy | 2 | 0.1-0.2 | 0.3-0.6 | 1-2 | 3-6 | 10-20 | % x d |
| Short term position accuracy | 3 | 200 | 600 | 1800 | 6000 | 18000 | m/15min |
| Instantaneous ground velocity accuracy | 4 | 0.4 | 1.3 | 4 | 13 | 40 | m/s |
| Instantaneous route accuracy | 5 | 0.07-0.2 | 0.2-0.7 | 0.7-2 | 2-7 | 7-20 | degree |
| Instantaneous heading accuracy | 6 | 0.1 | 0.3 | 1 | 2 | 7 | degree |
| Instantaneous attitude accuracy | 7 | 0.05 | 0.1 | 0.3 | 1 | 3 | degree |

INERTIAL NAVIGATION SYSTEM COST PERFORMANCE EFFECT

FORMULA PROPOSED

$$C/Co = [(Nm/ho)/(Nm/h)]^X$$

- (Nm/ho) = 3 Nm/h = long term position accuracy of the reference system
- (Nm/h) = Long term position accuracy of the unknown cost system
- Co = Reference system cost
- C = Cost of the considered system
- X = Coefficient 0.30 < X < 0.35

Interval of validity $0.30 \leq (Nm/h) \leq 30$

NAVIGATION - ACCOUNTING FOR COST

DERRICK J HAMLIN BSc., MSc
GEC Avionics Limited
Rochester, Kent, ME1 2XX. UK

SUMMARY

The ideal navigation system, from the point of view of the supplier, is a saleable system. From the user's point of view, the ideal navigation system would support the operational functions of his vehicle with perfect precision at all times. That the matching of these aims has been challenging is evidenced by the staggering array of systems which have been marketed for the purpose since the 1940's, the range of principles employed, their characteristics and their costs.

It is the purpose of this paper to survey the contending technologies for the navigation element of various types of guidance and control applications and to illustrate the practical constraints on both supplier and user in attaining the ideals.

The paper provides examples from both civil and military fields which are relevant to guidance and control, where improvements in cost effectiveness are being achieved.

1 INTRODUCTION

\$100,000 for a single electronic unit designed merely to tell you where you are may well be surprising to those who are unfamiliar with the use of Inertial Navigation Systems; the admission that such devices are expected to be wrong by perhaps 5km in normal operation is usually treated with disbelief; and the realisation that many civil aircraft boast three each, increases the mystery.

At the heart of this puzzle, which has taxed guidance system manufacturers for more than thirty years, is the need in many diverse applications for autonomy of operation. Radio navigation receivers are far more cost effective by comparison, but they are dependent on external influences which cannot be regarded as trustworthy in all circumstances.

The traditional users of navigation systems - ships and aircraft - have generally been content to suffer the cost penalties associated with their purchases provided that the units function reliably and reasonably accurately.

As weapons have become more autonomous, however, these too have begun to demand high precision over long operational flight-times. Unlike aircraft though, where the cost of the navigation sensing might be only one fiftieth of the cost of the air vehicle, the contribution can rise to one fifth in the case of missiles, unmanned vehicles and land vehicles when significant periods of autonomous operation are required. (Reference 1 and 2).

Three conflicting demands then are placed on the navigation system architecture. It must operate (everywhere and at all times) with high precision, require no external aids, and cost a relatively insignificant amount compared with the vehicle.

The diversity of customer choice, however, sometimes makes selection of a suitable device an onerous task. The problems are no less for the guidance system supplier who perpetually strives to match the laws of physics to the three criteria - autonomy of operation, accuracy and cost.

The object of this paper is to give some of the reasons for the conflict between these factors and to show the broad relationships between the factors for some of the presently available systems.

It concentrates on the position determination element of guidance and excludes consideration of target seekers, command links, and flight control. It also attempts to make the subject more manageable by avoiding the various combinations of integrated system, even though these can contribute greatly to the accuracy and robustness of navigation performance in some applications.

2 THE NAVIGATION REQUIREMENTS

It is instructive, first of all, to describe the navigation function which is separated into two parts for this purpose - an active part and a passive part. The active process, that of locating the vehicle, refers to setting the vehicle in position and on a desired track at a particular time. It is closely related to flight guidance, energy management, minimum flight separation, terrain avoidance, 4-D rendezvous and mission management.

The passive process, that of localisation, refers to detecting the position of the vehicle at a given time. In conjunction with accurate flight-times, radio communications,

and perhaps electro-optical or radar sensors, this function is closely related to target acquisition, weapon aiming, gun-laying, and air traffic control, for example. In addition, in many cases it supplies wind vector estimation, vehicle rotation rates and precise attitude, heading and velocities from the same system. It is therefore truly a multifunction facility, which accounts for its potentially high value in supporting the flight and mission avionics of a manned aircraft, but it also accounts, to a great extent, for its relatively high cost.

Of primary importance to the overall navigation system cost is the accuracy required by the user and this is naturally dictated by his application. Over a wide range of applications, performance needs may be anything from 1 metre at one end to 10 thousand metres at the other.

An accuracy of 1 metre is required for surveys and one system can be lowered 6 km down an oil-well drill shaft while reading out three dimensional position to this accuracy. Further down the scale, the navigation community encompasses, for instance, the road navigation of public service vehicles, ballistic missile guidance, military vehicles, submersibles and ships - essentially in a decreasing order of accuracy.

The lower end of performance is adequate for long-range civil airliners and, by way of example, for interplanetary missions. (The Pioneer spacecraft was well within this accuracy after its 800 million kilometer leg to Jupiter.)

Autonomy is the the next cost driver. Does the system have to operate within the specified performance limits for long periods without help or can its errors be corrected occasionally? Can it receive reliable radio signals or is it underground, below the sea surface or in a built-up area? More particularly, is the user sensitive to temporary or permanent loss of external radio aids as in the case of pilots of long-range civil aircraft or battle-field commanders?

Not surprisingly, the wishes of the customer are reflected in the price he pays; and prices range over more than two orders of magnitude. Even the lowest cost of radio navigator at about \$2,000 is expensive for some users, and its range of operation quite limited. At the other end, \$500,000, the costs reflect more the desire either for accurate, unaided operation over periods of many hours or for extremely high accuracy over much shorter periods. Moreover, these costs relate to discrete equipments, not necessarily to integrated systems. (In this document, costs have been normalised to 1984 U.S. Dollars using an approximate annual inflation rate of 7% where appropriate).

3 THE PRINCIPAL SYSTEMS

Trimming back the number of navigation technologies to a manageable selection for this analysis is probably unjust to those omitted members of the community which have perfectly valid application in certain circumstances. However, the following list illustrates the major contending techniques and encompasses the general principles of all the others.

RADIO NAVIGATION AIDS
Short Range Navigation, Area Navigation
Loran-C
Omega
GPS Navstar

AUTONOMOUS NAVIGATION SYSTEMS
Inertial (Platform)
Inertial (Strapdown)
Doppler/Heading

COMMUNITY NAVIGATION SYSTEMS

A broad grouping of types leads to a distinction between non-autonomous systems based on radio transmission, fully self-contained or autonomous systems, and systems based on data-links which either remotely measure the vehicle's position or relate each member of a community to the others. Clearly, the last mentioned category is not autonomous and depends on the maintenance of meaningful transmissions between members. When data-links are employed for navigation it is generally due to their adoption for quite different operational reasons. Nevertheless, if they can be used, a marked reduction in vehicle navigation cost is possible. Typical examples are unmanned target aircraft and remotely piloted vehicles but, as far as the general user is concerned, the restricted line-of-sight operational radius impairs the value of the technique. JTIDS, an example of this group, is primarily a communication system but its correlation processing in principle permits triangulation among its users. Accurately computing positional information from the ever-changing geometry of a group of vehicles presents a non-trivial problem however and maintenance of continuous, reliable communication for this purpose makes its cost effectiveness questionable if used purely as a navigational aid for some of the members of the community.

Concentrating now on the radio navigation aids, in Table 1, VHF area navigation systems have been selected in order to illustrate the cost and performance of short range receivers with coverage up to about 200 km for air vehicles. They form the lowest

cost navigation systems with prices ranging from \$2,000 for general aviation VOR receivers alone, to \$5,000 for VOR/DME fits and to \$25,000 for the most sensitive VOR/DME-based full area navigation systems. (Reference 3). In addition to their limited range, however, they suffer from interference and anomalous propagation effects causing errors of the order of 3 km at medium ranges. (Reference 4 and 5). There is little chance of reducing the cost or the errors of these systems because they are produced in relatively small quantities - being restricted in application largely to the aviation world - and they are essentially 'analogue' in nature and unable to benefit from the increasingly advantageous prices of digital components.

The range of operation of Loran-C is far more attractive and can extend to 2000 km over water but propagation anomalies cause errors to increase from 50 metres for pure ground-wave propagation, to 500 metres or more when corrupted by sky waves. Costs for commercial equipments range from \$1,000 to \$20,000 and military equipment specifications push the price beyond \$30,000. Here again, there is generally little attraction in reducing the costs, due to the limited appeal of a relatively local aid, although the accuracy in good conditions is valuable for some applications such as off-shore operations and a \$1000 unit is now available for marine use. (Reference 6).

Omega is a significant improvement over Loran-C in terms of uniform global coverage and the redundancy inherent in the multiple Omega and VLF transmitters provides protection for the user against spurious functional failure in individual transmitters. Its errors, though, are much greater than Loran-C due to the ionospheric, wave-guide propagation mode on which Omega depends. The best modelling available for the systematic propagation errors results in errors of the order of 3 km to 5 km but less complete models are implemented in some cases and higher errors are possible in certain circumstances. Multi-frequency processing, transmitter selection logic, and propagation modelling have made heavy demands on the computer architecture in the past but this has been rectified in more recent designs by introducing microprocessors and lower cost store technology. Major improvements have occurred in the size and cost effectiveness of Omega receivers and the facility would be more attractive but for the promises of the Global Positioning System - Navstar.

Without doubt, the prospect of automatic position determination to an accuracy of 10 or 15 metres, anywhere on the surface of the earth, is the most profound development in the field of navigation in many years. (Reference 7 and 8). That this can be achieved in the majority of applications at a cost to the user comparable with the far less capable conventional long and short range radio aids has to be regarded as a useful contribution to the cost effectiveness of guidance technology.

The user equipment costs shown in Table 1 are projections from earlier programme cost targets but these are now gradually being vindicated. As the production quantities increase it is probable that an increased dependence on VLSI will occur and cost will be reduced accordingly. This is due to the largely digital nature of the techniques involved and relative ease of standardising the circuitry in silicon, compared with the previous systems. At the more expensive military end, however, the price of the system is inflated by the need to guard the overall system performance by careful integration with an inertial navigation system. A further, large cost increment is added in the case of high performance aircraft due to the need for controlled receiver beam-patterns that demand complex antenna installation and beam-forming processes.

These aspects draw attention again to the critical problem of user autonomy. The user system can be jammed in certain limited circumstances, the geographically fixed ground elements can be destroyed, and, more importantly, the satellites in principle can be disabled or simply switched off. Potential users in the civil aviation field are reluctant to trust its aid if it can be instantly removed without warning, and field commanders are suspicious of the total dependence of their forces on a single, even slightly weak link. On the other hand, Transit has been in faithful service to both the military and civil worlds for many years and it clearly illustrates that GPS, with its significantly greater capability, will be even more widely used - subject to the approval of the tax-payers of the United States of America.

Therein lies a further difficulty for non-American users. Installation and support costs for radio navigation aids are extremely high and Table 2 lists some estimates for these costs over a twenty year period for comparison with the user equipment cost. (Reference 9). For the purposes of illustration, it is assumed that there is a community of 50,000 users on the North Atlantic, Eastern Seaboard Loran-C chain. The twenty year transmitter non-recurring and support costs per user equipment would correspond to the equivalent of \$7,000, or \$350 per year for each user.

For Omega, the 'hidden' cost would be \$18,000 over the twenty years and for GPS-Navstar it amounts to \$130,000 per user assuming 50,000 global users. There is, at present, no intention of charging potential users for their use of these systems.

4 AUTONOMOUS DEAD RECKONING SYSTEMS

In the case of velocity vector dead reckoning and inertial navigation systems, the situation is quite different. Their autonomy and freedom from interference is assured but, for the user, the cost of good performance is high.

The two practical forms of velocity vector navigation, shown schematically in Figure 1, comprise airdata-derived velocity or doppler radar coupled with a compatible heading source. Airspeed is corrupted by the errors in the estimate of windspeed while doppler measures true groundspeed and has a velocity accuracy which is as good as an aircraft inertial system. The primary cost limitation to both, however, is the heading measurement required to resolve the groundspeed vector into geographical axes. For complementary performance in each case, the heading component cost is greater than that of the velocity sensor. Note that there is a single integrator in each channel which transforms North and East velocities into position.

Performance of the airdata-based system is sometimes adequate for short periods of a few minutes, and it makes a reasonable supporter for Omega and the short range radio aids. In addition, cost, including a gyromagnetic heading source can be lower than the \$3,000 given in Table 3 provided that the navigator or pilot is interpreting the information. At the higher cost end, the navigation function is generally only an adjunct to other uses of high quality air data but, in either case, the determination of the wind vector is by far the dominant performance factor.

With a one degree gyromagnetic heading sensor, a doppler solution is probably feasible for significantly less than \$35,000 but, it seems, none is available for air-vehicles requiring a general navigation capability.

Matching the attitude and heading reference system performance to that of the doppler causes a major shift in cost but good system combinations are now becoming available with more successfully harmonised characteristics.

5 INERTIAL SYSTEMS

The acceleration vector, or inertial, navigation principle may be represented generally by the diagram shown in Figure 2. The diagram shows across the centre, the vehicle's acceleration measurements being corrected and resolved into earth's axes before being integrated to give velocity and then position. Earth-related-compensation terms are computed from the inertial system's own measurements of velocity and position in order to take into account movement over a spherical, rotating earth. Gravity compensation removes the fixed bias in the vertical direction.

The vehicle's acceleration vector is resolved into earth-related axes either by ensuring that the package of accelerometers, which forms an orthogonal set, remains stably oriented with respect to the earth's surface, or the accelerometer directions are followed very accurately as the vehicle manoeuvres. The first approach is a gimballed platform and this ensures that the sensitive sensor elements are isolated from the worst effects of the vehicle's manoeuvres. However, the cost of the platform-supporting equipment is high due to the high manufacturing accuracies required and the difficulties of achieving high reliability of electro-mechanical components.

The second approach, the strapdown inertial navigator, deletes the gimbals and their attendant problems but introduces almost impossible challenges for the design of the sensors, particularly the gyros. (Reference 10). In short, it turns out that the resulting strapdown cost for an aircraft INS is very similar to that of an equivalent gimballed version but the equipment reliability can be improved by two to five times with corresponding improvements in support costs.

The two integrators are clearly fundamental to acceleration vector navigation and they likewise have a fundamental effect on system performance. In the first place, advantageously, they attenuate short term sensor noise and produce very smooth measurements of position in comparison with doppler which uses only one integrator. The second effect of the integrators is highly detrimental. They accentuate sensor bias and drift errors as the square and cube-law of time and need extremely good, expensive sensors to bound the position errors to useable levels.

A typical system drift characteristic for gyro errors of the order of $1^{\circ}/\text{hr}$ might be represented in Figure 3 where, over 1000 seconds, the position error follows a third order law to about 8 kilometers. Its attraction over short flight intervals, however, is the slow start to the third order error characteristic.

This is reflected in the performance/cost relationship for various system types given in Table 4, (Reference 11) with short range missile grades providing, in the main, flight guidance stabilisation rather than true navigation.

Increasingly, AHRS units are being configured to provide navigation outputs as well as other vehicle parameters and, in this case, the drift errors tend to range from 30km/hr to 500km/hr when used in isolation. These errors may seem large but systems of this sort are usually velocity-aided with airdata or doppler information.

At the more accurate end, medium and high grade inertial navigation systems seem to range in price from \$80,000 or \$100,000 to maybe \$500,000 but the values in this case are more difficult to pin-down. Aircraft inertial navigation systems have ranged in performance from about 0.4km/hr to 10km/hr but in the main lie between one and two km/hr with velocity errors of the order of 0.5m/sec. Ship systems are more massive devices with more accurate sensors that require a benign environment for stable performance.

A breakdown of costs as shown in Table 5 for a typical gimballed INS would show a dominant contribution from the sensor cluster and gimbals. The computing element tends to be far less significant. The rest of the components, with the exception of the control and display unit, would be common to many applications and these rough proportions would probably be maintained in most cases.

6 EXAMPLES OF COST REDUCING METHODS

Various methods of cost reduction have been implemented in navigation systems, the most successful being in the area of digital techniques and in the adoption of large scale integrated circuitry. For example, one approach to improved doppler design has achieved marked reductions in doppler sensing costs for aircraft by means of simplified antenna manufacture, and by eliminating microwave and frequency tracker hardware. Figure 4 illustrates the unit.

More dramatic savings can be made in limited applications by precisely matching the capability of the sensor to the environment. The single axis doppler shown in Figure 5 costs less than \$1,000 to manufacture.

The manufacturing complexity of high performance strapdown systems is still a significant impact on their cost but units of the types shown in Figure 6, employing Laser gyros, have demonstrated an exceptional reliability in operational service - 7000 hrs mean time between failure in civil aircraft and over 2000 hrs in high performance military aircraft. These results were achieved immediately on introduction into service. The maintenance costs are also reduced by a factor of 8 over conventional platform systems due to elimination of periodic calibration and clean-room maintenance facilities, reduced spares handling, and improved self-monitoring.

Encouraging results can be achieved with inertial systems by simply leaving out some of the sensors as in the land-vehicle heading reference system illustrated in Figure 7 which is used in conjunction with an odometer (speed measuring equipment).

CONCLUSION

The paper attempts to relate the dominant navigation technologies to the performance and cost pressures experienced by users and suppliers. Although emphasizing a limited range of techniques, there are numerous variations on each and there are many others available for consideration in particular applications.

Some broad performance and cost bounds have been given which relate to the most promising and widely adopted systems and examples of successful improvements in the cost effectiveness of specific types of sensors have been presented.

| | Nominal Accuracy | | Cost Bracket (\$K) | |
|------------------|------------------|-------|--------------------|------|
| | Low | High | Low | High |
| VOR/DME, DME/DME | 3km | 0.5km | 5 | 25 |
| LORAN-C | 500m | 50m | 1 | 30 |
| OMEGA | 10km | 3km | 20 | 60 |
| GPS-NAVSTAR | 150m | 10m | 15 | 50 |

Table 1 Radio Navigation Systems

| | Non-Recurring Cost (\$M) | Estimated Support Cost (\$M) Over 20 Years |
|-------------|--------------------------|--|
| LORAN-C | 100 | 250 |
| OMEGA | 300 | 600 |
| GPS-NAVSTAR | 2500 | 4000 |

Table 2 Non-Recurring and Support Costs

| | Nominal Accuracy (% Distance Gone) | | Cost Bracket (\$K) | |
|---|------------------------------------|------|--------------------|------|
| | Low | High | Low | High |
| Air Data, Gyromagnetic Heading, Wind Vector | 10% | 4% | 3 | (25) |
| Doppler/AHRS | 2% | 0.3% | 35 | 70 |

Table 3 Dead Reckoning - Velocity Vector

| | Nominal Accuracy | | Cost Bracket (\$K) | |
|----------------------------------|------------------|----------|--------------------|------|
| | Low | High | Low | High |
| Short Range Weapon Grade (Agile) | 10m/s | 1m/s | 10 | 30 |
| Attitude and Heading Systems | 500km/hr | 30km/hr | 15 | 45 |
| Inertial Navigation Systems | 10km/hr | 0.4km/hr | 80 | 200 |
| Ships Inertial Systems | 10km | 0.5km | 200 | 500 |

Table 4 Inertial Navigation Systems

| | |
|----------------------------|-----|
| Inertial Measurement Unit | 45% |
| Platform Drive Electronics | 10% |
| Central Processor | 15% |
| System Interfacing | 15% |
| Power Conditioning Unit | 5% |
| Control and Display Unit | 10% |

Table 5 Approximate Cost Contributions in a Gimballed INS

REFERENCES

1. Hoag, David G., "Strategic Ballistic Missile Guidance - A Story of Ever Greater Accuracy"
Astronautics and Aeronautics, May 1978
2. Newbauer, John, "U.S. Cruise Missile Development"
Astronautics and Aeronautics, Sept 1979
3. Whitaker, Richard, "General Aviation Avionics"
Flight International, 12 Dec. 1981
4. Harlow, R. A., "Flight Measurements of Area Navigation System Performance Using Various Combinations of Ground Aids and Airborne Sensors"
IEEE Position, Location and Navigation Symposium, Dec. 1980
5. Johansen, Henry, "A Survey of General Coverage Nav aids for V/STOL Aircraft - A VOR/DME ERROR MODEL"
NASA CR-1588
6. Aviation Week and Space Technology,
p75, 5 Mar 1984
7. Jones, S.S.D., "GPS Navstar: A Review"
The Journal of Navigation, Vol. 32, No. 3, Sept. 1979
8. Henderson, Col Donald W., and Strada, Lt Cdr Joseph A., "Navstar Field Test Results"
The Institute of Navigation National Aerospace Symposium, 6-8 March 1979
9. Van Etten, J.P., "Medium Accuracy, Low-Cost Navigation: Loran-C Versus the Alternatives"
30th Technical Meeting of the Avionics Panel, AGARD
10. Peterson, Roland, "Advantages of Gimballed Inertial Navigation Systems".
NAECON '76 RECORD
11. Wu, Dr Thomas K., Filiatreau, Major Thomas R., and Gilmore, Jerold P., "A Design Concept for Low Cost Inertial Guidance of Tactical Weapons"
NAECON '78 RECORD

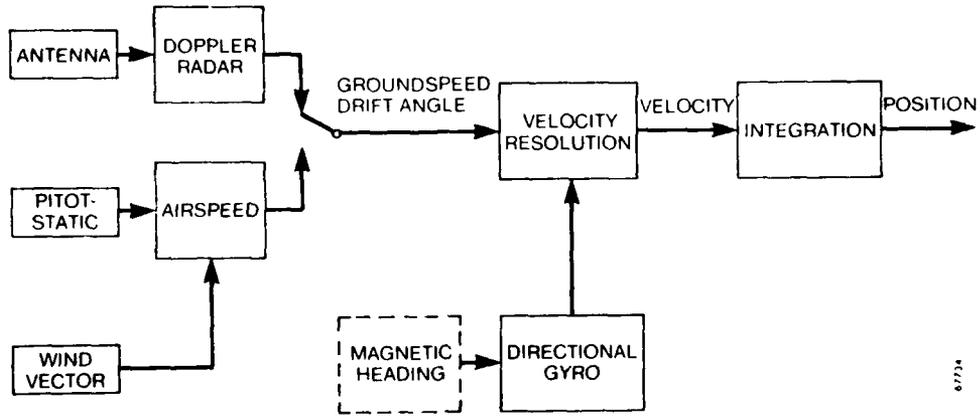


Figure 1 Velocity Vector Navigation

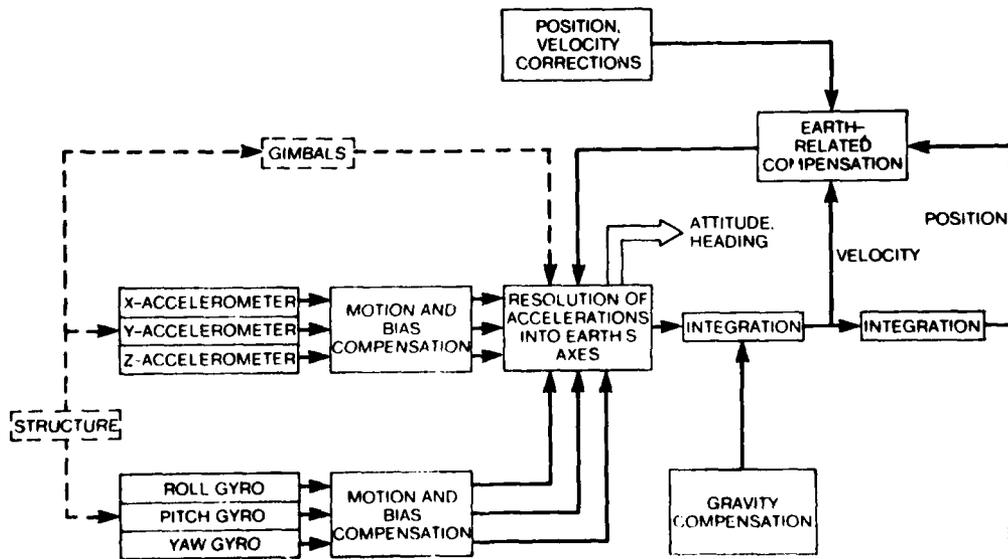


Figure 2 Acceleration Vector Navigation

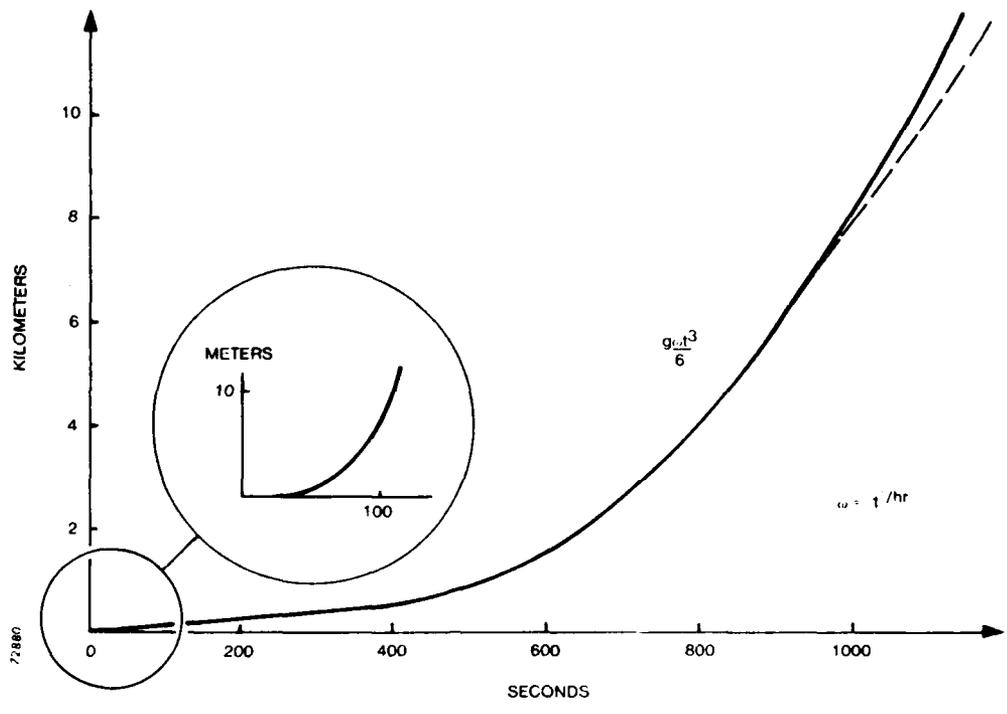


Figure 3 Inertial Error Characteristic

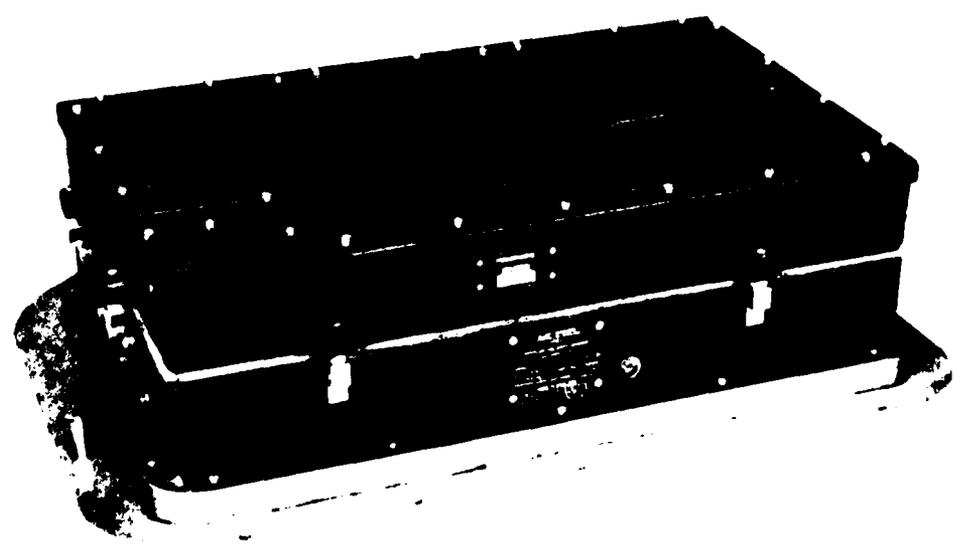


Figure 4 AD 660 Doppler Velocity Sensor

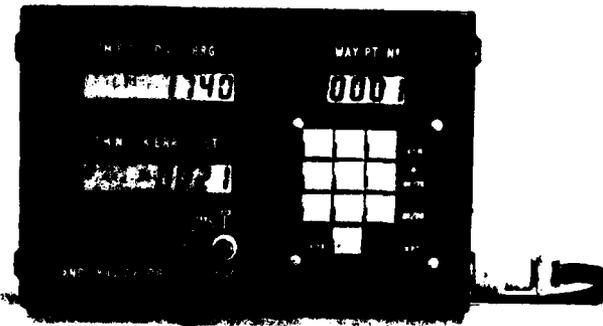


Figure 5 AA 5501 Vehicle Doppler Speed Sensor

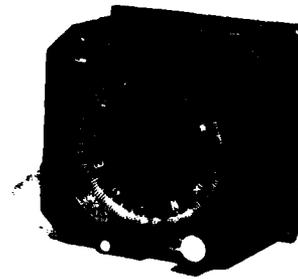


Figure 6 H-421 Laser Inertial Navigation System
(Courtesy Honeywell Inc.)

**NAVIGATION
CONTROL UNIT**



**VEHICLE
MOTION SENSOR UNIT**



**ANALOGUE
DISPLAY UNIT**

Figure 7 Land Navigation System

DESIGN ADEQUACY: AN EFFECTIVENESS FACTOR

by

Dr A.R.Habayeh
 Naval Air Systems Command
 Code AIR-320H
 Washington D.C. 20361, USA

The fundamental concept of system effectiveness underlies the decision making process of many people throughout history. The concept of system effectiveness will be reviewed and examined from the perspective of weapon systems consisting of launch platforms, targeting avionics, weapons, and targets. The tie-in between the various elements of a weapon system will be discussed. System effectiveness methodology provides insights into the design, development, utilization and tradeoff evaluation of diverse types of systems. Because of the multiplicity and uncertainty of the variables involved, the system effectiveness framework used in this paper is probabilistic. The application of system effectiveness to hardware systems is based on three effectiveness factors: (1) reliability (dependability), (2) readiness (availability), and (3) design adequacy (capability). The paper deals with design adequacy as a key system effectiveness factor. Design adequacy is a measure of how well a system performs its functions. It is the most desired factor in the definition, design, and early stages of system development. The paper presents a design adequacy quantification methodology and shows the relationship between design limitation and adequacy. The design adequacy methodology is based on the measures of adequacy, system parameters, subsystem parameters and the employment phases of the system. In a weapon system context, the performance parameters of a guidance and control subsystem, are interdependent with the parameters of the remaining subsystems. The paper deals with three employment phases of a weapon system. The three phases are: (1) prelaunch phase, (2) free flight phase, and (3) end-game phase. Examples based on air-to-air missiles are given to illustrate these relationships and concepts.

1.0 Introduction

New and emergent systems, like at no time before, must be responsive to increased demands for: more and better performance; lower life cycle cost; longer operating life; more productivity; and real improvements and attendant greater efficiency, in the planning and allocation of fiscal resources. System effectiveness analyses, conducted as early as possible, and continued throughout a system's lifetime, are a sound means of achieving this goal. In the early design, research, development, test and evaluation period, system effectiveness analysis will provide the needed tradeoff analyses of alternatives and options, and methodologies for achieving objectives and goals, and approaches for decision and choice making. This also provides needed insight into the dynamics and critical design parameters so important in the assessment of Design Adequacy, the primary factor in performance achievement.

In recent times, the design and development of systems have become more complex and have increased the need for systematic techniques, principles, and quantification methodologies. System designers have been faced simultaneously with demands to lower the cost and increase the performance under challenging conditions. The outcome of these demands is a tradeoff between performance and cost. Performance requirements invariably include severe reaction and response time constraints which cannot be met without the proper integration and resource allocation of personnel, intelligence, hardware and procedures. At the same time system development cost reductions, accelerated schedules, and lack of complete system tests prior to operational usage have combined to reduce the presence of adequate operational data banks, either in quality or quantity. Accordingly, it is necessary to have a comprehensive methodology of system analysis utilizing all available information both to pinpoint problem areas and to provide a numerical estimate of system effectiveness. The purpose of this paper is to put into perspective all the factors which must be considered in measuring the effectiveness of hardware systems, and to investigate the particularly important factor of Design Adequacy.

2.0 System Effectiveness Framework

Effectiveness is a desired outcome, state, consequence, or operation. Effectiveness is doing the right things right, to achieve the objective. System effectiveness is a measure of the extent to which a system can be expected to achieve a set of specific goals. It can also be thought of as the probability, P_{se} , that the system will achieve its objective. System effectiveness is an interdisciplinary subject which is based on probability, set theory, logic and uses the tools of system engineering, analysis and synthesis.

In general, hardware systems are designed and developed to meet specific objectives and goals; therefore, one needs tools and methods to determine how well they can achieve these objectives and goals. The need for methods to measure the achievement of design objectives invariably leads one to a concept of system

effectiveness and limitation. An accountable system effectiveness analysis is the most appropriate method for tradeoff analyses of alternatives, comparing evolutionary product improvement of existing systems to new optimized systems, evaluating the operational utility of competing options, and for the assessment of technology impact on systems. It further provides an analytical means for determining the viability of how often and how far one should pursue product improvement or incremental changes. The system effectiveness accountable factors are those which have significant influence on the desired outcome. For a hardware system, these factors are, performance, availability, modes of operation, organization, safety, survivability, environment, maintenance, repair, efficiency, compatibility and the like. Each factor in turn is dependent on a set of parameters such as performance parameters, parts count, failure rate, environmental conditions, real estate, down time, operating time, and similar variables.

System effectiveness deals with the behavior of the system throughout its life cycle. The quantification of system effectiveness must account for system limitations due to degradation in system readiness, physical failures, and design limitation. It follows that system effectiveness is a function of readiness, reliability and design adequacy (performance). A framework for system effectiveness can be based on the probabilistic notion of these three factors, and

$$P_{se} = P_{sr} P_r P_{da} \quad (1)$$

where, P_{se} is the probability that the system is effective. P_{sr} (system readiness) is the probability that the total system is operating, or ready to be operated satisfactorily. System readiness is a measure of the system support functions, e.g., maintenance, repair, and logistic. Readiness is the state of operability of the system given it is available. P_r (reliability) is the probability that the system will operate under stated conditions without malfunction for the duration of the mission; it is the mission reliability of the system. P_{da} (design adequacy) is the probability that the system will successfully accomplish its function given that it is operated within design specifications. It is the measure for performance capability. It reflects the degree of correspondence between specified performance and the capability actually demanded by the operational environment. The use of a system outside its design specification results in a limitation which reflects design inadequacy.

The system effectiveness factors in equation (1) are related to the factors used by the Task Group II of the Weapon System Effectiveness Industry Advisory Committee (WSEIAC). In the final report of January 1965, the Task Group II adopted the framework for system effectiveness evaluation based on three factors: (1) availability, (2) dependability and (3) capability.

The WSEIAC model is based on an enumeration of the significant system states over the entire mission. System states are discernible conditions of the system which result from events occurring prior to and during the mission. For example, the condition in which the system is available and operable is one state. The condition in which the system is completely or partially inoperable due to hardware, personnel, or procedure is another state. A system can transition from state to state during a mission. The structure of the model consists of three distinct parts, an availability row matrix [A], a dependability square matrix [D], and a capability column matrix [C]. It follows that system effectiveness [E] is

$$[E] = [A] [D] [C]$$

$$= [a_1 \ a_2 \ \dots \ a_n] \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & \dots & \dots & d_{2n} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ d_{n1} & \dots & \dots & d_{nn} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \cdot \\ \cdot \\ \cdot \\ c_n \end{bmatrix} \quad (2)$$

This model gives system effectiveness as a matrix of specified figures of merit or measures-of-effectiveness (MOE). The array of state probabilities at the beginning of a mission gives the availability matrix.

If the system has only two possible states: (1) operative, or (2) not operative (in repair); then the availability matrix "A" consists of two elements a_1 and a_2 ,

$$A = (a_1 \ a_2) \quad (3)$$

where, a_1 is the probability that the system is operable at a random point in time. a_2 is the probability that the system is in repair at a random point in time. The probability that the system is in the first state is,

$$a_1 = \frac{\text{Mean-time-between failure (MTBF)}}{\text{MTBF} + \text{Mean-time-to-repair (MTR)}} = P_{nr} \quad (4)$$

$$\text{MTBF} + \text{Mean-time-to-repair (MTR)}$$

The probability that the system is unavailable, i.e., it is in the second state is,

$$a_2 = \frac{MTTR}{MTBF + MTTR} = 1 - p_{sr}$$

(5)

If the system is available and there is no repair during the mission, then both models equation (1) and (2) are equivalent. Under this condition, p_{sr} is the single shot probability of success and $p_{sr} = a_1$. Since the system is available and repair is not possible during the mission, the dependability matrix reduces to one element, d_{11} , where d_{11} is the probability that the system is operable at the end of the mission, given that it was operable at the start of the mission, which is equal to the mission reliability, p_r . Similarly, the capability matrix contains one element, $c_1 = P_{da}$.

2.1 Measures-of-Effectiveness (MOEs)

A measure-of-effectiveness (MOE) is any index which indicates the quality of a system. In the simplest case it may be a measured physical quantity, such as velocity, range, payload of an aircraft, or the number of targets that can be tracked by a sensor. On the other hand, it may be a calculated quantity based on measurement, such as mean down time, or mean-time-to-repair (MTTR). Also it can be a predicted quantity based on measurement and/or simulation. For example, the probability that a system will survive an attack will require prediction because there will be some uncertainty about the attack condition and environment. MOE's serve to indicate what can be expected from the system; they must be in an operationally oriented form that can be readily understood and used. They should reflect the essence of the system; i.e., the reason behind its existence.

System effectiveness is the term often used to describe the overall capability of a system to accomplish its mission. For a missile, system effectiveness is the probability that the missile will kill the target when called upon to do so under specified conditions. Therefore, the single shot kill probability, p_{kss} is the missile system effectiveness, $p_{se} = p_{kss}$. By this definition, the missile fails (i.e., does not kill) (1) if it is unavailable or inoperable when needed or, (2) if it is operable when needed but fails to kill. The first case accounts for system physical failure and degradation in readiness. The second case accounts for design limitation of the weapon system. The phrase specified conditions implies that system effectiveness must be defined in terms of the requirements placed upon the system, indicating that failure to kill and conditions of use are related.

2.2 System Partitioning and Hierarchy

The process of advancing from a coarse analysis to a fine analysis of a set of logical possibilities is achieved by means of partitioning. Similarly, the process of determining what should be in a subsystem, in an electronic package or on a solid state wafer is a partitioning problem. The objective of system partitioning is to optimize functional and structural interaction of the system's building blocks with respect to design objectives. A given system can be partitioned into different functional levels according to topology or function, or both. The selected partitioning levels are based on complexity of function and structure. The following complexity levels are considered here: systems, subsystem, functional block, function block, circuit, and element, as shown in Figure 1.

The vertical structure of the hierarchy consists of a multiplicity of substructures nested within the system. The horizontal structure of the hierarchy is characterized by mutual support, interaction, or interdependence among the elements at each level. Achievement of higher level objectives depends upon the functioning of the lower level structures. At each level in the hierarchy, system functions are dependent on subordinate functions. The relationship between functions is a chain interaction whereby changes in system functions at one level will impact functions at other levels. The higher the functions in the hierarchy, the more inclusive they are. That is, system functions at one level account for a multiplicity of other functions in the lower levels. Consequently, there is a hierarchy of MOE's.

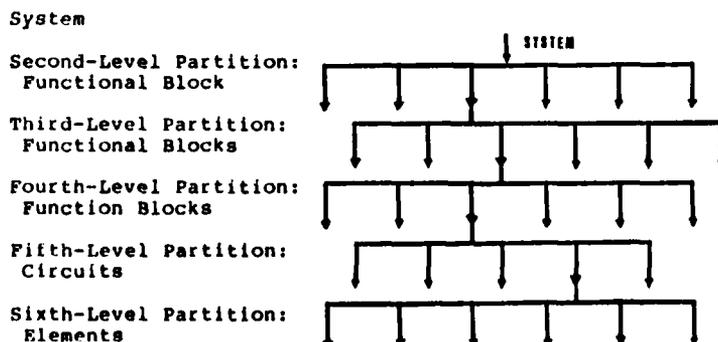


Figure 1. Partitioning Levels of a Hardware System

Combat effectiveness is the term often used to describe the overall capability of a force to accomplish its objective. From a system analysis point of view, a hierarchy of combat exists with vertical and horizontal structure as shown in Figure 2. The upper levels (1, 2, 3, and 4) represent operational combat, and the lower levels (5, 6, and 7) represent system and technology.

In this paper the MOEs of the different levels in the hierarchy of combat are:

| <u>Level</u> | <u>MOE</u> |
|-------------------------|--------------------------|
| 1. War | Resources depletion rate |
| 2. Warfare Mission Type | Draw-down |
| 3. Battle | Kill rate |
| 4. Duel | Exchange ratio |
| 5. Weapon system | Single-shot-kill |
| 6. Subsystem | Figure of merit |
| 7. Technology | Figure of goodness |

THE HIERARCHY OF COMBAT

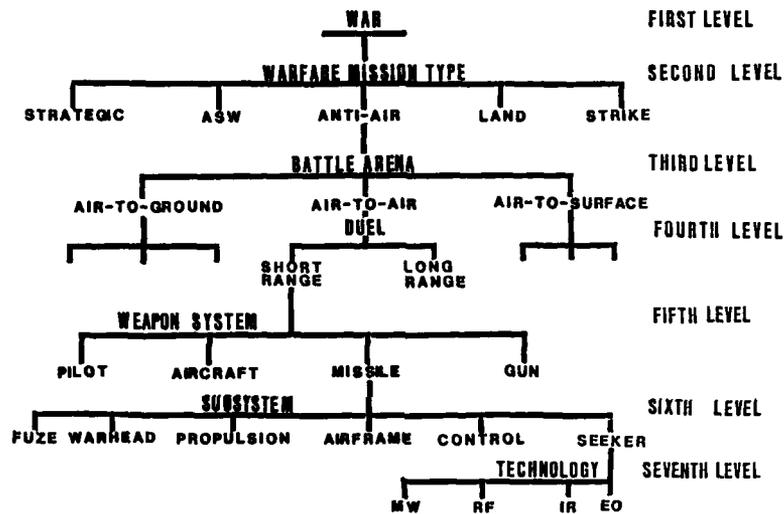


Figure 2. Structure Of The Hierarchy Of Combat

The weapon system level is the link between the upper and lower levels. Its effectiveness measure is the single-shot-kill probability, P_{kss} . The duel level deals with the relationship between the parameters at the operational and system levels, and its MOE is the exchange ratio. At the subsystem level, MOEs are generic performance outputs of more than one technology, such as miss distance, off-axis angle, kinematics, range capability, and lethality. They are referred to as figures of merit (FOMs), or measures of performance (MOPs). For example, FOMs for an electro-optical (EO) seeker are tracking rate, aspect capability, and minimum resolvable temperature (MRT), which accounts for the sensitivity, resolution, signal to noise ratio, the modulation transfer function (MTF), and the field of view (FOV) of the various building blocks of the seeker.

The initial step in analyzing combat effectiveness is to determine the level for which effectiveness is required. Under some circumstances it may be necessary to quantify effectiveness at lower levels in order to establish the impact on higher levels. For example, the impact of advanced technology on the operational combat levels cannot be quantified directly. One needs a middle ground to provide continuity in the evaluation, e.g., the weapon system level. Improvement in effectiveness at the weapon system level, i.e., P_{kss} , is then projected into the next higher level, i.e., exchange ratio.

3.0 Effectiveness and Limitation

A system attains its objectives not through a single operational action, but rather through a series of fundamentally different ones. Therefore, system objectives are attained through an integrated series of individual or combination of events, and it makes no difference whether all these actions are performed by one building block,

different subsystems or groups of them. Primary interest lies in determining what these events consist of, when, and why they occur. An event is defined in terms of subset and sample space of an experiment, system, or organization. A sample space of a system is a complete set of elements such that any state of the system corresponds to exactly one subset or element in the set. In this context, the elements of the sample space of a system are the binary combinations of the events associated with its functions. For example, if the functions of a system are a, b, c, then the system sample space consists of $2^3 = 8$ elements. The elements are the binary combinations of the events associated with the functions, e.g., AB, AC, BC, ABC.

System effectiveness may be made tractable by the application of system limitation. This is represented by the set of events for which the desired outcome is not achieved. Probabilistically, limitation events are the complement of effectiveness events. Equation (6) gives the relationship between system effectiveness, limitation, and the three effectiveness factors;

$$P_{se} = P_{sr}P_{da}P_r = 1 - P_{sl} \quad (6)$$

where, P_{sl} = system limitation.

The relationship between system effectiveness and limitation is illustrated in the system space diagram, Figure 3, where the rectangle represents the total system space. This is the set of events associated with the system outcomes. It includes effectiveness and limitation events. In this diagram, a point in the rectangle represents a possible system situation resulting from a possible trial. The limitation region in Figure 3 accounts for all the causes of system limitation. Sometimes in the conceptual phase of system development, it is easier to identify limitation than effectiveness. That is, determining what the system cannot do is easier than determining what it can do. It should be observed that due to multiplicity and uncertainty of the variables, system effectiveness is neither absolute nor deterministic, i.e., it cannot be measured exactly by an instrument, but it is rather a probabilistic concept and it should be so treated.

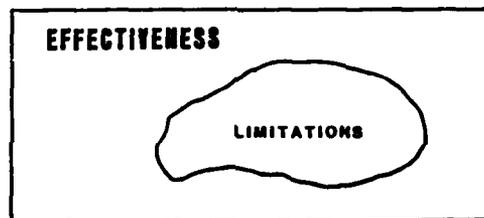


Figure 3. System Space Relating System Limitation and Effectiveness

Probabilities are defined in terms of events. It follows that the analytical framework of system effectiveness can be represented in terms of the events of the three effectiveness factors as shown in the Euler diagram of Figure 4. In the diagram, the limitation regions are labeled unreadiness (unavailability), inadequacy (incapability), and unreliability (undependability). Total system limitation is represented by the Venn diagram in Figure 4, which is the area bounded by the limitation regions. System effectiveness is the region of the rectangle which is outside the Venn diagram. Figure 5 is another representation of system effectiveness and limitation, it is the complement of Figure 4, where system effectiveness is the inner region of the Venn diagram, and system limitation is the remaining part of the rectangle. The Euler diagrams of Figure 4 and 5 show that the limitation and effectiveness events overlap each other. Both diagrams show eight disjoint regions, seven of them represent limitation events and one represents effectiveness. The limitation regions in Figure 4 are not mutually exclusive and therefore the limitation events of unreadiness, inadequacy and unreliability may or may not be statistically independent. Therefore, the effectiveness factors are interdependent operationally, and system effectiveness requires they occur simultaneously.

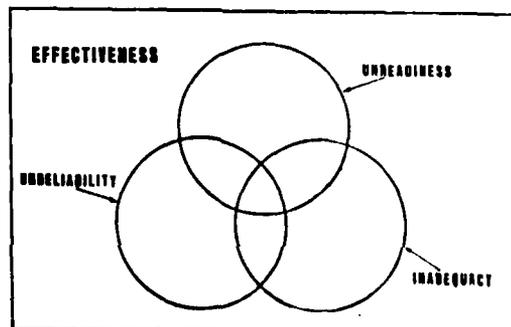


Figure 4. Events Diagram Of System Effectiveness And Limitation

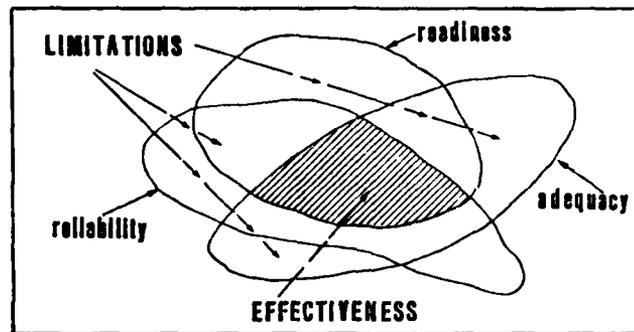


Figure 5. Events Diagram Of System Limitation And Effectiveness

If the events in Figure 4 and 5 are not statistically independent, then the the probability P_{gr} , P_{da} , and P_r are conditional probabilities. From an operational point of view, readiness or availability events must occur first, i.e., the system must be ready or available at the time of mission. Therefore, availability events are not conditional on the reliability or design adequacy events.

Inadequacy, unreliability and unreadiness events are random in nature and sequential conditioning is not a reasonable representation of their occurrence. For example, unreliability events may occur during testing, storage, transportation, or captive carry, and if undetected it will influence the new state of design adequacy. On the other hand, design inadequacy can be the state of the system before its application. Consequently, inadequacy events had occurred prior to system employment regardless of reliability and readiness. Design adequacy is built into the system a priori and is fixed; what changes in the adequacy perspective is the environment, e.g., new applications, conditions, or threat. Similarly, reliability is built in the system at the time of putting it together. Degradation in reliability is a function of time and the environment. Readiness represents procedural, e.g., repair, maintenance and operational, e.g., logistic, training, and captive carry events.

Quantification of the system effectiveness factors is done in steps at different times. Design adequacy is evaluated in the early stages of system design and development where, readiness and reliability data are not available. It is determined from performance analysis by simulation or experimentally. Consequently, design adequacy is quantified by assuming that readiness and reliability events had occurred as specified. Reliability is assessed from the block diagram of the system, and using parts count and failure rate of the hardware. Reliability calculation assumes that the performance and readiness are as specified. On the other hand readiness is estimated from the reliability assessment, and the mean-time-to-repair (MTTR). The estimation assumes that the system will operate within its design limits. Some aspects of reliability, (e.g., redundancy, hardware quality and testing) or readiness (e.g., modular design, and human factors) can be accounted for in the early stages of developing design adequacy. It follows that, system effectiveness quantification is an interactive process. After the system has been operational, feedback can be used to update the quantification of design adequacy, reliability and readiness. In general, the three effectiveness factors are quantified separately from each others. Accordingly it is reasonable to assume that the effectiveness events are statistically independent.

4.0 Design Adequacy:

Design adequacy is an indicator of how well a system performs. It is a measure of the capability of a system to perform satisfactorily in its operational environment. System design adequacy is the probability that a system will successfully accomplish its function, given that it is ready and reliable. This probability is a function of such variables as the operational environment, system accuracy, design limits, interaction and organization of its subsystems, system inputs, and the influence of the operator. System design inadequacy accounts for the usage of a system beyond its design capabilities, even though the reliability and system readiness may approach ideal values. For example, the ineffectiveness of a missile designed to engage subsonic targets when used against supersonic targets is due to design inadequacy even if $P_r = P_{gr} = 1.0$. Design adequacy reflects the degree of correspondence between specified performance and the system capability demanded by the operational environment.

Simulation, experimentation, and system analysis provide the tools of quantifying design adequacy. The design adequacy of a system, P_{da} , can be estimated from knowledge of the system performance when it is operated within specification, and from knowledge of the performance required to achieve success in the mission for which effectiveness is being evaluated. To clarify this statement, consider an airborne communication

system with a mission duration of 5 hours. Further, consider that on the average during 10% of the mission time, the aircraft will be called upon to operate at a range beyond the performance capabilities of the system, even though the communication system operates at optimum performance level, i.e., within design specifications. In this example, $P_{da} = .9$, which implies that the system is adequate for the mission only 90% of the time. Assume that the reliability of the system, $P_r = .95$ and that the system will operate within specification for a period of 5 hours. Furthermore, an estimate of system readiness, P_{sr} , based on operating time, down time, free time, and available spares shows this probability to be 94%. Therefore,

$$P_{se} = (.95) (.9) (.94) = .8037$$

If the hardware reliability were improved so that $P_r = .999$, and system readiness were $P_{sr} = .97$, then

$$P_{se} = (.999) (.9) (.97) = .87$$

Similarly, for an air-to-air missile, if for the majority of combat encounters the missile does not intercept the target on the average 10% of the time due to target maneuverability, then the design adequacy of the missile is $P_{da} = .9$.

4.1 System Design Adequacy Model

The quantification of system design adequacy, P_{da} , is a difficult and complex system problem. This is because the design capability of a system is dependent on the performance of its subsystems, and the operational environment. Therefore, the system performance parameters are operationally dependent on the subsystem design parameters. Consequently, the quantification of system design adequacy requires the determination of the statistical and the functional relationships between the parameters. For example, the range capability of a missile as a system parameter depends functionally on the propulsion, airframe, and the seeker subsystem design parameters. Similarly, the range capability of a radar as a system parameter depends operationally on the antenna, transmitter, receiver, and the display subsystem design parameters. Accordingly, the system design limitation due to one system parameter is influenced by other system parameters. An approach to render tractability to design adequacy is the application of system design limitation P_{dl} , where,

$$P_{dl} = 1 - P_{da} \tag{7}$$

Equation (7) defines the events of design limitation as complementary to the events of design adequacy. The quantification of the design limitation requires the identification of the factors which affect the accomplishment of the mission. The next step in the process is the determination of the functional and the statistical relationship between the system parameters. That is - determine if parameter "a" is dependent functionally and statistically on parameters "b" but is independent of parameter "c".

Consider system (S) with two system parameters "a" and "b"; let "A" and "B" be the design limitation events associated with the parameters "a," and "b" respectively. The system parameters are interdependent and their influences on the design limitation of the system are not mutually exclusive. That is, system design limitation due to "A" will partially occur during design limitation due to B. The example is illustrated by the Euler diagram of Figure 6.

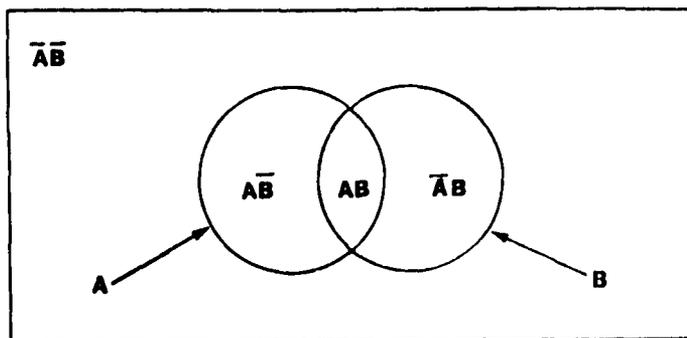


Figure 6. Euler Diagram Of Two Not Mutually Exclusive System Events

The rectangle of Figure 6 represents the total system design space i.e., design adequacy and design limitation. In the diagram the system design limitation events are represented by the two overlapping circles, where \bar{A} is the complement of A. The system design adequacy is defined by the region of the rectangle which is outside the Venn diagram. The design limitation due to either/or events A and B is

$$P_{d1} = P(A + B) = P(A) + P(B) - P(AB) \quad (8)$$

If the events A and B are statistically independent, then by definition, $P(AB) = P(A)P(B)$. Similarly, for a case involving three parameters "a," "b," and "c," one can write the probability of design limitation to be

$$P_{d1} = P(A+B+C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC) \quad (9)$$

as shown in Figure 7.

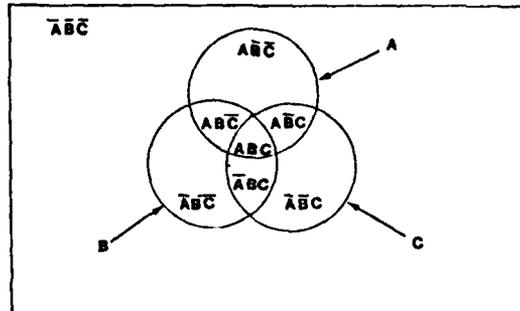


Figure 7. Euler Diagram Of Three Not Mutually Exclusive System Events A, B, C

Design adequacy can be calculated directly from the Euler diagram describing the total relationship between adequacy and limitation. Consider the case of Figure 6, design adequacy is the region outside the Venn diagram and represents situations in the which both A and B occur:

$$P_{da} = P(\bar{A}\bar{B}) = 1 - P(A) - P(B) + P(AB) \quad (10)$$

For the case of three not mutually exclusive events represented by Figure 7

$$P_{da} = P(\bar{A}\bar{B}\bar{C}) = 1 - P(A) - P(B) - P(C) + P(AB) + P(AC) + P(BC) - P(ABC) \quad (11)$$

For the general case of "n" not mutually exclusive events representing "n" system parameters: a_1, a_2, \dots, a_n :

$$\begin{aligned} P_{da} &= P(\bar{A}_1\bar{A}_2\bar{A}_3\dots\bar{A}_n) \\ &= 1 - P(A_1) - P(A_2) - \dots - P(A_n) + P(A_1A_2) + P(A_1A_3) + \dots + P(A_1A_n) + P(A_2A_3) + \dots + P(A_2A_n) - P(A_1A_2A_3) - P(A_1A_2A_4) - \dots - P(A_2A_3A_n) + \dots + P(A_1A_2\dots A_n) \\ &= 1 - \sum_{i=1}^n P(A_i) + \sum_{i \neq j}^n P(A_iA_j) - \sum_{i \neq j \neq k}^n P(A_iA_jA_k) + \dots + (-1)^n P(A_1A_2\dots A_n) \quad (12) \end{aligned}$$

If the events $A_1, A_2, A_3, \dots, A_n$ are statistically independent, then the compound probabilities in (12) are the products of the probabilities of the events, i.e., $P(A_1A_2A_3) = P(A_1)P(A_2)P(A_3)$.

The preceding discussion is best illustrated by an example. Consider a tactical air-to-air guided missile (AAM) and the six system parameters and their functional relationships: (1) Off-boresight Angle (OBA); (2) Target Aspect (TA); (3) Missile Range; (4) Accuracy; (5) Lethality; and (6) Maneuverability.

Off-Boresight Angle (OBA) is the angle between the bore-sight line or armament datum line (ADL) and the line-of-sight (LOS) to the target. Angle-of-attack (AOA or α) is the angle between the ADL and the flight path vector \vec{V} . The relationship between (OBA), (ADL), (LOS), (AOA), velocity vector (\vec{V}), seeker gimbal angle and tip-off angle are shown in figure 8. Assume ejection launch, the missile will separate parallel to the ADL, and because of the flow field and differential velocity, it will tip-off and become closely parallel to the flight path of the launch platform.

Target aspect is the target profile presented to the seeker during lock-on, tracking, and the end game and is vitally important to a successful intercept. It is the prime indicator of the maneuvering intensity required to achieve an intercept, e.g., tail-on will result in slower closing speed (V_C) whereas head-on V_C is the sum of the target's and missile's individual speeds resulting in very high-gain maneuvering constants at intercept terminus. Target aspect effects the amount and quality of the missile seeker's tracking medium, e.g., exhaust heat for a heat seeking missile; and reflected RF signal for a radar missile. Various target aspects are shown in figure 8.

Missile Range, expressed as R_{min} or R_{max} , is functionally related to seeker sensitivity and its off-boresight tracking capability (seeker gimbal angle limits); aerodynamic profile requirements (maneuverability and closing velocity V_C); and propulsion subsystem performance. For every missile design there is a maximum aerodynamic range that is the physical resultant of the propulsion subsystem. Launch aircraft airspeed adds vectorally to the missile's airspeed thereby providing an additional range component. Target aspect and the resultant V_C can add or subtract to the actual maximum launch range. For example, in a classic forward quarter target intercept, the launch aircraft's weapons system may compute first firing at a target distance that exceeds the aerodynamic range of the missile, but is valid due to the combined target/missile closing velocity (V_C) which acts to cause intercept at a distance short of the missile's maximum aerodynamic range.

Accuracy is functionally dependent upon intercept kinematics plus seeker and fuze sensitivity and resolution. The missile's guidance and control subsystem and its propulsion subsystem must provide target intercept within a distance sufficiently close so that the fuze will cause a lethal warhead detonation.

Lethality is functionally attributed to accuracy, intercept kinematics, and warhead.

Maneuverability is the missile's ability to roll about its longitudinal axis, to pitch (pull g's) about its lateral axis and yaw about its vertical axis. It is a kinematic design parameter functionally attributed to aerodynamic (skid-to-turn or bank-to-turn control laws) and propulsion subsystems. Maneuverability is limited by aerodynamic and structural design adequacy.

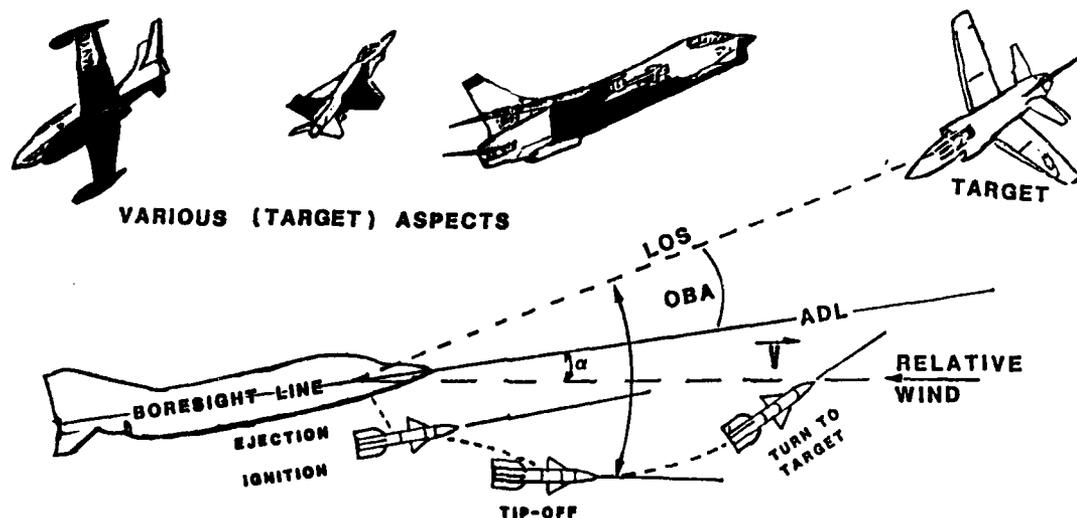


Figure 8. Relationship between OBA, ADL, LOS, \vec{V} , α and Tip-Off Angle. Tip-off angle = $k(OBA + \alpha)$ where $1.0 \leq k \leq 3$ is an empirical number, and gimbal angle must be \geq than tip-off angle to maintain lock-on.

Consider the case of four operationally related variables whose events are not-mutually exclusive and are statistically independent. Let "A" represent off-boresight limitation, "B" is aspect limitation, "C" represents range limitation and "D" is accuracy limitation events. Figure 9 shows the design limitation diagram for this case, it shows $2^4 = 16$ disjoint regions.

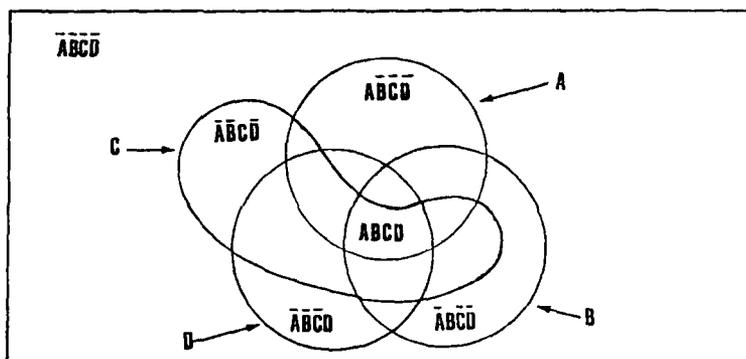


Figure 9. Design Adequacy/Limitation Diagram Of A Hypothetical Tactical Missile

Consider the case where, through simulation (using Monte Carlo or other techniques) of many encounters, the system design limitations are, $P(A) = .1$, $P(B) = .05$, $P(C) = .08$ and $P(D) = .05$. It is desired to know the design adequacy of this hypothetical missile. From equation (12),

$$P_{da} = P(\overline{ABCD})$$

Since A, B, C and D are statistically independent, then

$$\begin{aligned} P_{da} &= P(\overline{A}) P(\overline{B}) P(\overline{C}) P(\overline{D}) \\ &= (.9) (.95) (.92) (.95) = 0.75 \end{aligned}$$

For this example, $P(A)$ accounts for the limitations due to A only and limitations due to AB, AC, and AD, as shown in Figure 9.

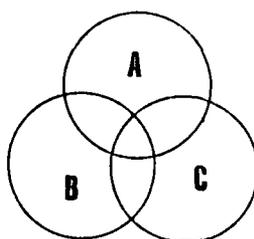
4.2 Design Adequacy Variables

If a system is operable when needed but fails to complete the assigned mission successfully, then the failure is associated with design inadequacy. As the operational stresses and the mission variability increase, the frequencies of physical failure and design limitation may also increase. Moreover, operational requirements sometimes exceed design objectives. For example, an increase in target maneuverability, or a decrease in vulnerability can result in a decrease of system effectiveness. Similarly, a surface-to-air radar designed for air targets may have no system effectiveness against land targets. Design limitation events occur very dynamically and in a mixed order during system employment. The outcomes of these events are represented by random variables.

Statistically, design adequacy is concerned with both discrete and continuous random variables. The number of design limitations (frequency) in a given mission, and the number of poorly designed parts, such as detectors with sensitivity limitation, power supply with limited current, seeker with limited gimbal angle, or kinematic parts of low performance are discrete random variables. Also, in air-to-air encounters, the time between successive gimbal angle limitation, the time between kinematic constraints and time between losing lock-on due to inadequate seeker sensitivity are continuous random variables. Similarly, the time between acceptable message transmissions in a communication system, or satisfactory delivery in a transportation system are continuous random variables. These examples suggest that design adequacy data can be collected in either form, as discrete or as continuous random variable observations, at the discretion of the data collection designer.

5.0 Design Adequacy Prediction of a Hypothetical Air-to-Air Weapon System

The system effectiveness analysis of an air-to-air weapon system involves consideration of five elements and their interactions. These elements are launch platform, pilot, avionics, missile, and target. Each element of an air-to-air weapon system consists of several subsystems. For example, the missile consists of five subsystems: guidance and control, airframe, propulsion, warhead, and fuze. A logical approach to the quantification of weapon system design adequacy is to consider three operationally sequential and interrelated phases. The three phases are (1) pre-launch, (2) intercept (free flight) and (3) end-game. Design adequacy of the weapon system is a function of the system limitation of each phase. System limitation probabilities during the three phases are: launch opportunity limitation P_{l01} , intercept limitation P_{i1} and kill limitation P_{k1} . Figure 10 is the design limitation diagram of the three phases of the weapon system.



A = (pre-launch limitations)
 B = (intercept limitations)
 C = (end-game limitations)

Figure 10. Design Limitation Diagram Of An Air-To-Air Weapon System

From Figure 10:

$$P_{d1} = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC) \quad (13)$$

Where $P(A) = P_{101}$, $P(B) = P_{i1}$, and $P(C) = P_{k1}$.

The system limitation events of one phase are not mutually exclusive of the limitation events of the other phases, and

$$P_{da} = (1 - P_{101}) (1 - P_{i1}) (1 - P_{k1}) = P_{10} P_i P_k \quad (14)$$

The probabilities P_{10} , P_i and P_k are conditional probabilities of the sequential events, pre-launch, intercept and end game, where P_{10} = probability of launch opportunity, P_i = probability of intercept, and P_k = probability of end-game kill.

P_{10} is the probability of being in a position within the Launch Acceptability Region (LAR) at some instant from acquisition time to the end of the encounter. It is the probability of in-envelope launch given a ready and reliable system. This parameter is a measure of the capability to shoot early, i.e., a high P_{10} is indicative of a high capability to shoot first, and is a function of target acquisition by the pilot and the weapon. It is a measure of the pilot's ability to maximize the probability of being within the missile launch envelope after target acquisition and before firing.

P_i = probability of intercept, given in-envelope launch. P_i is the probability of the weapon intercepting the target within the Intercept Acceptability Region (IAR), and accounts for the kinematic and guidance and control parameters.

P_k = probability of kill given intercept. P_k is the end-game probability of kill and is directly related to the missile intercept angle, the region of fuze activation, and warhead effectiveness.

Referring to the system effectiveness equation $P_{se} = P_{da} P_r P_{sr}$, and substituting for P_{da} from equation (14) gives the system effectiveness of the air-to-air weapon system expressed in terms of the conditional probabilities P_{10} , P_i , P_k , P_r and P_{sr} :

$$P_{se} = P_{da} P_r P_{sr} = P_{10} P_i P_k P_r P_{sr} \quad (15)$$

$$= (1 - P_{101}) (1 - P_{i1}) (1 - P_{k1}) (1 - P_{r1}) (1 - P_{sr1})$$

$$= P(\overline{A}\overline{B}\overline{C}\overline{D}\overline{E})$$

$$= P(\overline{A}/\overline{B}\overline{E}) P(\overline{B}/\overline{A}\overline{D}\overline{E}) P(\overline{C}/\overline{A}\overline{B}\overline{D}\overline{E}) P(\overline{D}/\overline{E}) P(\overline{E})$$

as shown in Figure 11, which shows there are $(2)^5 = 32$ disjoint regions.

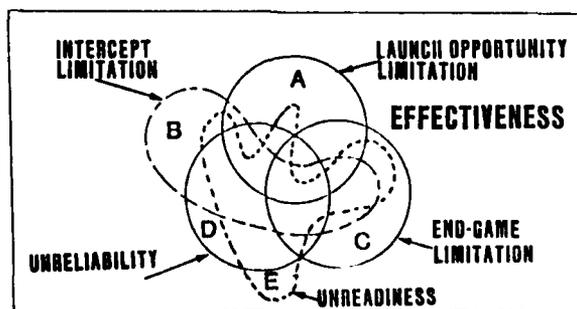


Figure 11. System Effectiveness/Limitation Diagram Of An Air-To-Air Weapon System

The launch acceptability region (LAR) is a volume representing the missile launch envelope within the target acquisition space. It is the region of valid launch and/or firing opportunities. The intercept acceptability region (IAR) is a volume within which the missile will intercept the target. The kill acceptability region (KAR) is a volume about the target within which given missile passage the weapon is lethal. The events, valid launch, intercept and kill occur in the LAR, IAR, and KAR regions. These events are multivariates and require three variables for their description.

5.1 System and Subsystem Parameters

The prime objective of system effectiveness analysis is the minimization of limitations; for the hypothetical tactical missile, limitation is the bounded area within the non-overlapping edges of the regions in Figure 11. The first step in the minimization process is to identify and quantify the sources of limitation. The next step is to establish the relationship between them. The third step requires relating system and subsystem parameters to the measures of design adequacy (MOA).

It is desired to relate the weapon system measures of adequacy P_{10} , P_i , P_k , to the system and subsystem functional parameters. Over the three phases of the engagement, the subsystems are (1) target, (2) pilot target acquisition avionics, (3) missile guidance and control, (4) missile airframe and propulsion, and (5) missile warhead and fuze. The system capability is the collective attribution of the performance of these subsystems. The operational weapon system parameters are: (1) target acquisition (pilot, avionics, and weapon), (2) off-boresight angle (OBA), (3) target aspect (TA) which is the target view presented to the launch aircraft or missile from any point in a target-centered sphere of view. All-aspect would be the capability to detect and track a target regardless of its presented aspect, (4) kinematics of platform, target and missile, (5) intercept miss-distance, (6) lethality, (7) physical characteristics. The functional relationship between subsystems parameters as the independent variables and system parameters as the dependent variables is given in Table 1. For example:

Pilot target acquisition = F (target characteristics, acquisition aids, guidance, aircraft and missile kinematics)

Off-boresight = F (target characteristics, pilot acquisition aids, guidance and control, kinematics)

Target Aspect Angle = F (target characteristics, seeker detection and tracking capability)

The weapon system Measures-of-Adequacy (MOA) as the dependent variables are related as follows:

- P_{10} = F (OBA, TAA, target acquisition)
- P_i = F (intercept miss-distance, kinematic, physical parameters)
- P_k = F (lethality, kinematic, physical parameters)

TABLE 1

| SYSTEM PARAMETERS | SYSTEM AND SUBSYSTEM PARAMETERS, LOCK-ON BEFORE LAUNCH AIR-TO-AIR WEAPON SYSTEM | | | | | | |
|---|---|--|-------------------------------|---|--|---|---|
| SUBSYSTEM PARAMETERS | TARGET ACQUISITION PILOT, WEAPON | OFF-BORESIGHT OFF-AXIS AT LAUNCH | TARGET ASPECT | KINEMATICS AIRCRAFT, TARGET MISSILE | MISS DISTANCE INTERCEPT | LETHALITY | PHYSICAL CHARACTERISTICS WEIGHT SIZE SHAPE |
| 1 TARGET CHARACTERISTICS MANEUVERABILITY VELOCITY, SIZE ALTITUDE RANGE RADIATION CROSS SECTION SIGNATURE | HIGHLY DEPENDENT ON THE TARGET CHARACTERISTICS | HIGHLY DEPENDENT ON THE TARGET | RADIATION CROSS SECTION GLINT | MANEUVERABILITY VELOCITY, RANGE ALTITUDE | MANEUVERABILITY VELOCITY, RANGE | STRUCTURE VELOCITY | INDIRECT THROUGH PROPULSION AIRFRAME WARHEAD SIZING |
| 2 PILOT ACQUISITION AVIONICS AIDES ALTITUDE REACTION TIME HAND-OFF TO WEAPON STRESS/FATIGUE | <ul style="list-style-type: none"> • ACQUISITION • STRESS • A/C MANEUVERABILITY • ACQUISITION TRANSFER | HIGHLY DEPENDENT A/C MANEUVERABILITY, STRESS, VELOCITY, HAND-OFF TO WEAPON, REACTION TIME | NO | AIRCRAFT KINEMATICS | SOME | NO | NO |
| 3 GUIDANCE & CONTROL SEEKER SENSITIVITY RESOLUTION FIELD-OF-VIEW CIRCLAR ANGLE TRACKING RATE SLW RATE GUIDANCE LAW CONTROL SURFACE LIMIT | <ul style="list-style-type: none"> • SEEKER SENSITIVITY • SLW RATE • TRACKING RATE • DETECTION PROBABILITY • CM CAPABILITY | <ul style="list-style-type: none"> • CIRCLAR ANGLE • SLW RATE • TRACKING RATE • GUIDANCE LAW • CM • CONTROL SURFACES LIMIT | SENSITIVITY FOV | CONTROL SURFACES LIMIT | <ul style="list-style-type: none"> • SENSITIVITY • TRACKING RATE • GUIDANCE LAW • CONTROL SURFACES LIMIT • CM | INTERCEPT MISS DISTANCE | WEIGHT, SHAPE SIZE |
| 4 AIRFRAME AND PROPULSION LIFT & DRAG ALTITUDE SURFACE AREA VELOCITY MANEUVERABILITY ANGLE OF ATTACK • REDUCED ROLL • CONTROL SURFACE LIMIT C.G. LOCATION WARRAD & SIZE WARRAD TYP WARRAD WEIGHT WARRAD SENSITIVITY WARRAD ACCURACY | <ul style="list-style-type: none"> • Thrust • Velocity • Maneuverability • Thrust Profile • Altitude | <ul style="list-style-type: none"> • Range • Rate • Maneuverability • Altitude | NO | MISSILE KINEMATICS • Range • Rate • Velocity | <ul style="list-style-type: none"> • Range • Rate • Time to Intercept • Maneuverability | <ul style="list-style-type: none"> • Range • Rate • Velocity | WEIGHT, VOLUME OF PROPULSION, AND AIRFRAME |
| 5 WARRAD & SIZE WARRAD TYP WARRAD WEIGHT WARRAD SENSITIVITY WARRAD ACCURACY | NO | NO | NO | NO | WARRAD WEIGHT TYP WARRAD SENSITIVITY & ACCURACY | WARRAD TYP WARRAD SENSITIVITY & ACCURACY | WARRAD SIZE SHAPE |

For the short range air-to-air case, the probabilities P_{10} , P_i and P_k are determined by simulating a large number of encounters and accounting for the various target characteristics such as maneuverability, velocity and radiation cross section. The simulation is mechanized into three consecutive stages representing the operational phases of the system. The simulation runs of the second stage assume the events of P_{10} has occurred, and those of the end-game stage assume the events of P_{10} and P_i have occurred. Data from mock combat can be used to validate the simulated values of P_{10} .

Table 2 shows the functional relationship between the system parameters and the MOA's.

TABLE 2
MEASURES-OF-ADEQUACY AND SYSTEM PARAMETERS
A/A WEAPON SYSTEM

| MEASURE-OF-ADEQUACY | BEFORE LAUNCH P_{10} | INTERCEPT P_i | END-GAME P_k |
|---|---|---|--|
| SYSTEM PARAMETERS | LAR: LAUNCH ACCEPT-ABILITY Region | IAR: INTERCEPT ACCEPTABILITY REGION | KAR: KILL ACCEPT-ABILITY REGION |
| TARGET ACQUISITION | DEPENDENT ON | SOME | NO |
| REACTION TIME STRESS ACQUISITION AID ACQUISITION TRANSFER | <ul style="list-style-type: none"> o PILOT TARGET o ACQUISITION AID o STRESS o ACQUISITION PROB-ABILITY | <ul style="list-style-type: none"> o HANDS-OFF o ACQUISITION AID TO SEEKER o COUNTER MEASURE | |
| OFF-BORESIGHT ANGLE | HIGHLY DEPENDENT ON OBS | SOME OBS ANGLE AND LAUNCH ENVELOPE | NO |
| TARGET ASPECT 360 DEG SEEKER ACQUISITION | DEPENDENT AND DETERMINES OBS ANGLE LIMITS | SEEKER SENSITIVITY | SOME FUZE SENSITIVITY |
| INTERCEPT MISS DISTANCE | NO | HIGHLY DEPENDENT | HIGHLY DEPENDENT |
| KINEMATICS TARGET MANEUVERABILITY RANGE VELOCITY AND ALTITUDE AIRCRAFT VELOCITY, ALTITUDE MANEUVERABILITY MISSILE MANEUVERABILITY RMAX, RMIN TIME-TO-INTERCEPT | <ul style="list-style-type: none"> o HIGHLY DEPENDENT ON ALL THREE KINEMATIC PARAMETERS o LAUNCH ENVELOPE | HIGHLY DEPENDENT ON THE TARGET AND THE MISSILE KINEMATICS | <ul style="list-style-type: none"> o TARGET VELOCITY MANEUVERABILITY o MISSILE RMIN MANEUVERABILITY VELOCITY |
| LETHALITY WARHEAD AND FUZE MISS DISTANCE | NO | NO | HIGHLY DEPENDENT |
| PHYSICAL PARAMETERS WEIGHT SIZE SHAPE | NO | DEPENDENT RMAX, VELOCITY MANEUVERABILITY | WARHEAD WEIGHT, TYPE AND SHAPE |

Figure 12 represents the design limitation during the three weapon systems employment phases. The figure relate the system parameters to the measures-of-limitation, MOL. Refer to equation (15) and let:

$$P_{10} = .85, P_i = .9, P_k = .95, P_r = .89, P_{gr} = .96,$$

then

$$P_{se} = (.85) (.9) (.95) (.89) (.96) = 0.62$$

However if $P_r = P_{gr} = 1.0$ then

$$P_{se} = (.85) (.9) (.95) (1.0) (1.0) = .726$$

With the assumption $P_r = P_{gr} = 1.0$, the system effectiveness equation reduces to $P_{se} = P_{da}$, and the measure-of-adequacy MOA, becomes measure-of-effectiveness MOE. For an air-to-air weapon system, there are other measures-of-adequacy, such as exchange ratio, first firing, first kill and time-in envelope.

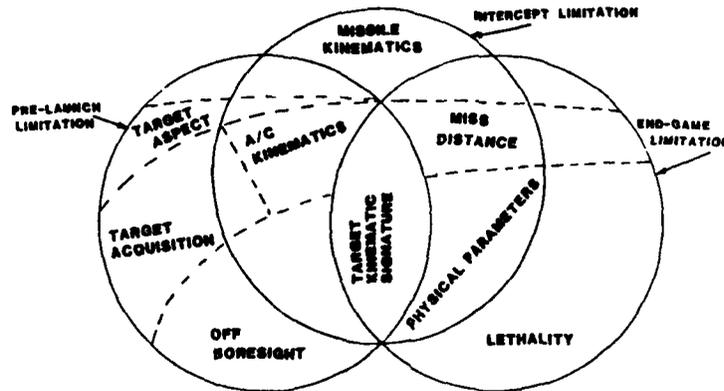


Figure 12. Design Limitation Diagram Of An Air-To-Air Weapon System

5.2 Design Adequacy Parameters Simulation

Design adequacy of an air-to-air missile includes sensitivity to variations in altitude, required G's or maneuverability of the launch platform, closing or opening range rate, and launch speed among others. Figure 13 shows kinematic sensitivity comparison between two missile designs. These plots summarize the results of several hundred Monte Carlo simulation runs based on a six degrees of freedom model. The kinematic parameters are altitude, target G's, range and closing velocity (V_c). The ordinate of the plots is the probability of intercept.

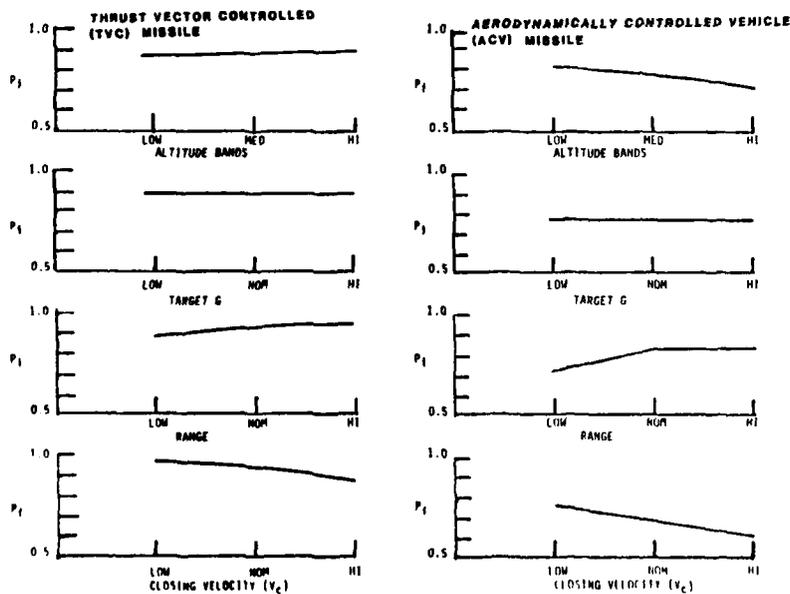


Figure 13. Probability Of Intercept P_i Versus Kinematic Parameters Of Two Missiles

Figures illustrate nominal "dogfight" missile envelope - target centered.

AAM LAUNCH ENVELOPES: 10 units altitude, coaltitude with target.

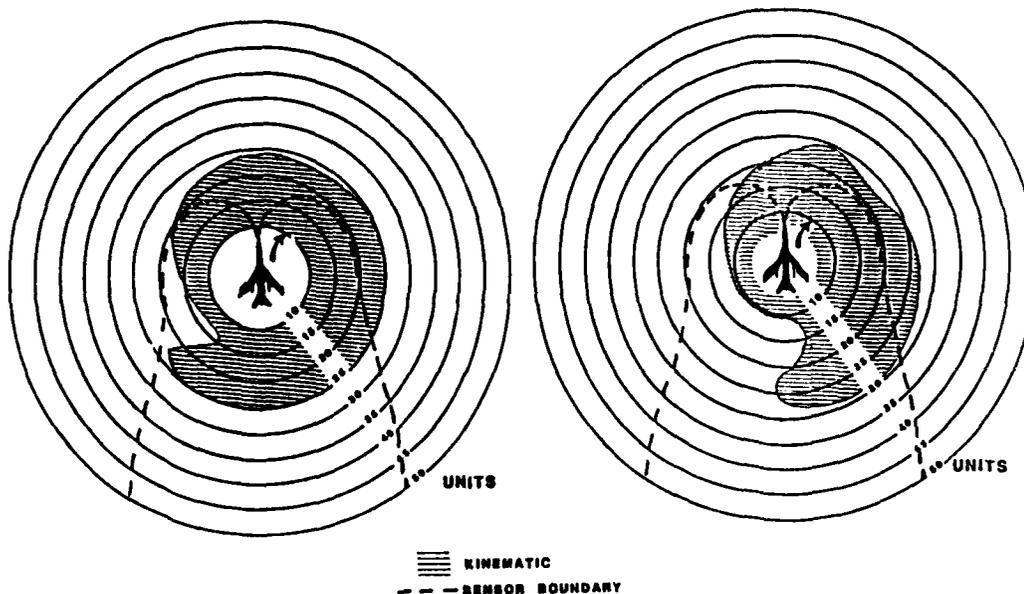


Figure 14.

Figure 15.

At launch: Ftr vel. (V_f) = $M0.8$ = Tgt vel.
Target starts immediate 5g horizontal
right-hand turn.

At launch: Ftr vel. (V_f) = $M0.8$, Tgt vel. (V_t) = $M0.95$
starts immediate 5g horizontal right-
hand turn.

The drama of design limitations is very apparent upon examination of the launch envelopes shown in Figures 14 and 15. The figures give system design boundaries due to kinematic factors (shaded areas) and sensor parameters (dashed area). Valid launch must occur within the overlapping regions of the two boundaries. One can observe that even a 0.15 Mach change in target velocity has a substantial effect on both inner and outer envelopes. It should be pointed out that these are instantaneous envelopes and are valid if and only if the target maintains the maneuvering condition indicated on the plots. That is, a sudden reversal in the direction of turn accomplished by a half roll will nullify a significant part of the envelope for launches inside the target turn beyond 10 units radius in Figure 15. The theoretical envelopes emphasize that operationally, the dynamics of the situation deform the launch envelopes in a highly unpredictable manner.

Conclusion

System design adequacy is an effectiveness factor which is a measure of the capability of a system to perform its function. It is the key effectiveness factor in the early stages of system development. Design adequacy of a system is a function of the performance parameters of its subsystems and the environment. The paper presents a design adequacy quantification model which accounts for the interdependence between the building blocks. In a weapon system context, the performance parameters of guidance and control subsystems are interdependent with the parameters of the remaining subsystems. Therefore, the effectiveness of the guidance and control subsystem is highly dependent on this interaction. The quantification methodology requires determining the MOA's, the system and subsystem parameters. The determination of accurate relationship between these parameters requires indepth knowledge of the hardware and is a very serious system engineering problem.

REFERENCES

1. Lanchester, F.W., "Aircraft in Warfare; the Dawn of the Fourth Arm", D. Appleton & Co., New York, 1918.
2. Weapon System Effectiveness Industry Advisory Committee (WSEIAC), AFSC-TR 6S-2 Vol. II, January 1965.
3. Reliability Engineering, ARINC Research, Prentice Hall, 1964.
4. Wohl, J.G., "System Operational Readiness and Equipment Dependability", IEEE Transaction on Reliability, Vol. R-15, No. 1, May 1966.
5. "System Analysis and Policy Planning", E.S. Quade and W.I. Boucher (ed.), American Elsevier Publishing Co., Inc., New York, 1969.
6. Habayeb, A.R., "System Decomposition, Partitioning and Integration for Microelectronics", IEEE Transaction on System Science and Cybernetics, July 1968.
7. Feller, W., "An Introduction to Probability Theory and Its Applications", Vol. I, John Wiley and Sons, Inc., 1962.
8. Parzen, E., "Modern Probability Theory and Its Application", John Wiley and Sons, Inc., 1960.
9. "Systems Engineering Methodology and Applications", A.P. Sage (ed.), IEEE Press, New York, 1977.

Acknowledgements:

Part of this paper was researched and written in 1982 when the author was at Yarmouk University, Irbid, Jordan.

THE EVOLUTIVE STATE: A NEW DEFINITION OF SYSTEM STATE TO INCLUDE SEQUENCING OF EVENTS

by

Prof. V. Amioia
Politecnico di Milano
Ist. di Elettrotecnica e Elettronica
Piazza Leonardo da Vinci 26
Milano, Italy

and

Ing. R. Somma
Selenia Spazio
Divisione Sistemi Spaziali
Via S. Alessandro 28/30
I-00131 Roma, Italy

ABSTRACT

The usual reliability modeling is based on the introduction of a system state (herewith referred to as macrostate) identified by the operating conditions (good, failed) of the system components. The inclusion of the information about the states sequencing, absent in the classical description, leads to the introduction of the concept of microstate. This concept leads to a transition diagram structure, whose peculiarity is exploited to derive a useful computational procedure for the state probabilities.

1. INTRODUCTION

The classical approaches to system reliability are based on the concept of system state, which is identified by the set of operating conditions of the system components.

If each component, as usual, is described by two mutually exclusive operating conditions, i.e. good and failed, a system composed of n components, i.e. of order n , has 2^n states.

In this paper, for reasons which will become evident later on, each one of the above states will be referred to as macrostate or operating state of the system.

Let us now consider the following simple example. A system of order 3 has the reliability block diagram of fig. 1.

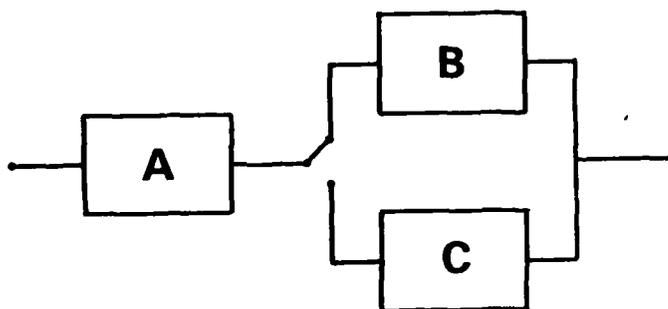


Fig. 1

B and C are two identical units, C being the cold redundancy of B. A represents the switch, which is needed to activate the redundancy in case of failure of B.

The following realistic assumption is made: the effect of a failure of A is that the switch remains in the position occupied at the time the failure happens.

Being the system of order 3, it has the 8 macrostates given in tab. 1, where 0 represents the "good" and 1 the "failed" condition.

TAB. 1

| n. | A | B | C | SYSTEM |
|----|---|---|---|--------|
| 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 1 | 0 |
| 2 | 0 | 1 | 0 | 0 |
| 3 | 0 | 1 | 1 | 1 |
| 4 | 1 | 0 | 0 | 0 |
| 5 | 1 | 0 | 1 | 0 |
| 6 | 1 | 1 | 0 | ? |
| 7 | 1 | 1 | 1 | 1 |

The graph which represents the transitions, in case of no repair, is given in fig. 2, where λ 's are the components failure rates.

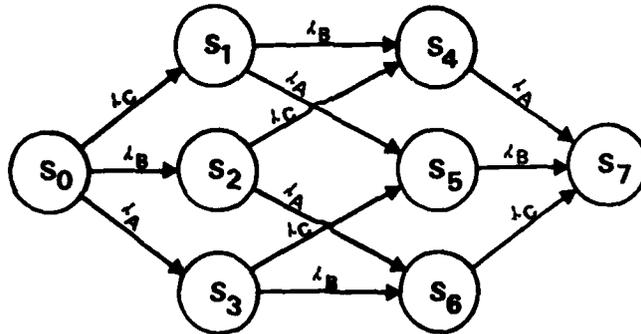


Fig. 2

In tab. 1, no information about the system ability to perform is associated to the macrostate S_6 , in fact this macrostate may be good or failed respectively if B fails before A or viceversa, i.e., according to the graph of fig. 2 if S_6 is reached through the path $S_0-S_2-S_6$ or $S_0-S_3-S_6$. Therefore, once the probability of S_6 has been evaluated by whatever a method, one cannot associate this probability to any of the system operating conditions, unless it is possible to split it into the two contributions pertinent to the two paths identified above.

Now, as in general such a partitioning is not possible a posteriori, there are cases, like the one used as an example, or in general when the needed information is carried out by sequencing of events leading to a certain state rather than by occupation of the state itself, where the classical description in terms of macrostates is not sufficient to solve the reliability problem.

All the above motivates the introduction of the microstate (or evolutive state) concept, which enriches the information content of the classical macrostate by adding the sequencing of events. The resulting peculiar system states transition diagram will be presented in this paper, together with some theoretical results allowing an easy evaluation of the evolutive states probabilities.

2. THE MICROSTATE AND THE DYNAMIC TREE

Microstate or evolutive state of a system is the set of information carried by a macrostate and by the sequence of transitions leading to it.

According to the above definition, each macrostate which in a classical reliability transition diagram can be reached by more than one path, will generate a different microstate for each different path.

With reference to the example of fig. 1, this is clarified in the following fig. 3:

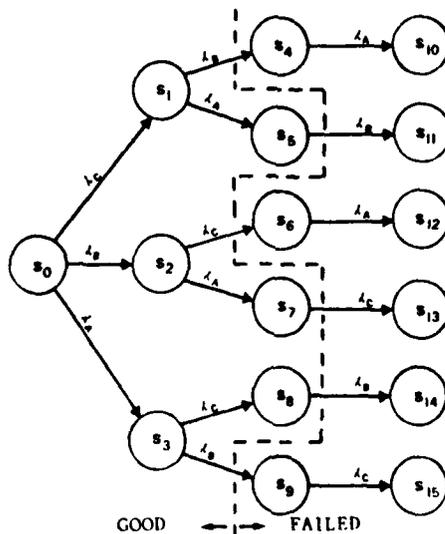


Fig. 3

In fig. 3, differently from fig. 2, each microstate, s , can be reached by only one path, therefore it evidences both the operating condition of the components and, what constitutes the added information, the sequence of events leading to such an operating condition.
 The comparison of fig. 2 and 3 puts the sets of microstates, S_i , and that of macrostates, s_i in the following correspondance:

$$\begin{array}{ll}
 S_0 \Leftrightarrow \{s_0\} & S_4 \Leftrightarrow \{s_3\} \\
 S_1 \Leftrightarrow \{s_1\} & S_5 \Leftrightarrow \{s_5, s_8\} \\
 S_2 \Leftrightarrow \{s_2\} & S_6 \Leftrightarrow \{s_7, s_9\} \\
 S_3 \Leftrightarrow \{s_4, s_6\} & S_7 \Leftrightarrow \{s_{10}, s_{11}, s_{12}, s_{13}, s_{14}, s_{15}\}
 \end{array}$$

The splitting of S_6 into s_7 and s_9 eliminates the ambiguity present in the description in terms of macrostate. In fact now, according to the description given previously, s_7 pertains to the set of good states, while s_9 is a failed state.
 The elimination of the above ambiguity, is paid in terms of enlarged number of system states, which is now greater than the usual 2^n (for a system of order n). In general the number of microstates is infinite, as can be immediately derived with reference to the simple repairable component with an unlimited repair possibility, the microstate transition diagram of which is depicted in fig. 4, as it is well known for those familiar with the Renewal theory:



Fig. 4

The problem is now the search of properties, algorithms and rules to cope with such a large number of microstates (and then of equations), so as to have the opportunity of exploiting to advantages of the enriched information content.
 To that purpose let us consider a system of order n (i.e. having 2^n macrostates); as each component has two mutually exclusive operating conditions, the most general case, from a transition point of view, is the one in which each component can be subjected to the transition leading it to the state opposite to the one presently occupied. Less general cases may be obtained by simply eliminating the branches corresponding to not allowed transitions.
 Let us assume that the system is in the microstate in which all its components are in the "good" condition and the system is released for the first time into operation.
 When one of the components changes its operating condition, the system changes its microstate reaching one of the n possible configurations characterized by 1 failed component and $(n-1)$ good components.
 Next event may be the repair of the failed component or the failure of one of the $(n-1)$ good components.
 Notice that in the macrostate description the repair is a backward transition leading the system into the macrostate occupied in advance. On the contrary, in microstates terms also the repair is a forward transition leading the system to a "new" microstate, characterized by the same macrostate information but by a different trajectory.
 Notice also that the microstate reached by the sequence "failure of component i , failure of component j " is different from the one reached by the sequence "failure of component j , failure of component i " even if, once again, the resulting macrostate is the same.
 Going on with the same procedure, one has immediately that starting from each microstate the system may, in general, reach n new microstates, some of which may pertain to macrostates already encountered during the system evolution.
 The resulting transition diagram is a directed n -ary tree [1], as shown in Fig. 5, where the figures on the branches identify the components which cause the transition

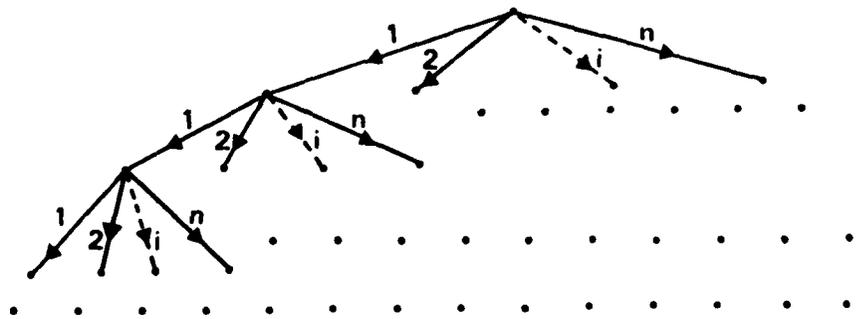


Fig. 5

This is a very well known topological structure [1], formed by nodes (the microstates) and directed branches (the transitions) connecting them.

The structure is organized by levels. The starting level (level 0) has only one node, called root, corresponding to the initial microstate, in which the system is put into operation for the first time.

Next level (level 1) has in general n nodes, which are reached starting from the root by the n directed branches, each representing the failure of one component.

Next level (level 2) contains microstates which require two subsequent events to be reached.

In general, the level of a microstate is defined by the number of transitions necessary to reach that microstate starting from the root, s_0 .

A collection of rules to define the set of the macrostates and that of the microstates and to pass from one set to the other is contained in [2].

If now the attention is put on the general microstate s_i and the assumption is made of constant transition rates, one can state that:

the system is in s_i at the time $(t + \Delta t)$ if it was already there at the time t and no transition occurred in the subsequent time interval Δt , or if the system was in the microstate s_j , immediately preceding s_i , at time t and the transition leading from s_j to s_i has taken place in the time interval Δt . In formula, and in probability terms:

$$P_i(t + \Delta t) = \{1 - (\sum \lambda)_i \Delta t\} P_i(t) + \lambda_{ij} P_j(t) \Delta t \quad (1)$$

from which one has immediately:

$$\dot{P}_i(t) = -(\sum \lambda)_i P_i(t) + \lambda_{ij} P_j(t) \quad (2)$$

where $(\sum \lambda)_i$ represents the sum of the transition rates of the branches going out from s_i .

In particular:

$$\dot{P}_0(t) = -(\sum \lambda)_0 P_0(t) \quad (3)$$

due to the fact that s_0 is the first microstate and has not a predecessor.

If, as usual, the initial condition is:

$$P_0(0) = 1 \quad (4)$$

the solution of eq. (3) is given by:

$$P_0(t) = e^{-(\sum \lambda)_0 t} \quad (5)$$

Now, each microstate of the first level, s_1 , derives from s_0 , therefore, according to (2) and considering (5):

$$\dot{P}_1(t) = -(\sum \lambda)_1 P_1(t) + \lambda_{1,0} e^{-(\sum \lambda)_0 t} \quad (6)$$

Solving each one of the equations (6) the probabilities of the microstates of the first level may be easily evaluated.

These probabilities are used to write down the equations of the microstates of the second level, which are similar to eq. (6). And so on for the subsequent levels.

All the above outlines that even if the number of variables, the probabilities of the microstates, has been increased with respect to the description in terms of macrostates, their evaluation is simpler, as by proceeding by levels each variable requires the solution of an uncoupled differential equation.

If now the attention is put to the structure of fig. 5, and to the procedure that have generated it, one can easily observe that starting from s_0 , each transition from a microstate partitions the subsequent microstates into separated possible system evolutions, which never convergn.

A basic theorem refers to the above separation property; it is given herebelow.

Theorem (fundamental):

let s_i be any microstate of a system of order n, and s_{ij} , $j = 1, 2, \dots, n$, the microstates which can be reached with only one transition from s_i (i.e. the sons of s_i). Each s_{ij} is in turn the root of a subtree.

The probability associated with the overall set of the microstates in the subtree of any root s_{ij} , is given by the product of the following two terms:

- the probability, $P_{F_{ij}}(t)$, associated with the overall set of microstates generated by s_{ij}
- the ratio between the transition rate pertinent to the transition from s_i to s_{ij} and the sum of all the transition rates going out of s_i (connecting then s_i with all the possible s_{ij}).

In formula, if $P_{(i_j + F_{i_j})}(t)$ indicates the probability of the set of microstates of the subtree of root s_{ij} , one has:

$$P_{(i_j + F_{i_j})}(t) = \frac{\lambda_{i_j, i}}{\sum_{j=1}^n \lambda_{i_j, i}} P_{F_i}(t) \quad (7)$$

Proof [3]:

The differential equation associated with any microstate s_{ij} is (see also eq. (2)):

$$\dot{P}_{i_j}(t) = \lambda_{i_j, i} P_i(t) - \sum_{k=1}^m \lambda_{k, i_j} P_{i_j}(t), \quad \forall 1 \leq j \leq n$$

where the summation in the second term represents the cumulative transition rate of the transitions leading the system out from s_{ij} .

Each term in the above summation appears, as a positive term, in the differential equation of the microstates which can be reached from s_{ij} with only one transition. This last equation contains also a negative summation of terms, which appear as positive terms in the subsequent equations and so on. If in such a process is continued until an absorbing microstate is reached, the summation disappears from its equation.

Therefore, when in the second term of the differential equation of any microstate a negative summation is present, the microstate is not absorbing and the elements of the summation appear as positive terms in the equation of the microstates generated by it.

As a consequence sum of the differential equations relative to s_{ij} and to the microstates originated by it with any possible number of transition is:

$$\dot{P}_{(i_j + F_{i_j})}(t) = \lambda_{i_j, i} P_i(t), \quad \forall 1 \leq j \leq n$$

Therefore by summing over j :

$$\dot{P}_{F_i}(t) = \sum_{j=1}^n \lambda_{i_j, i} P_i(t)$$

Furthermore, according to the previous equation:

$$P_i(t) = \frac{1}{\lambda_{i_j, i}} \dot{P}_{(i_j + F_{i_j})}(t)$$

By introducing this last equation in the previous one we obtain:

$$\dot{P}_{(i_j + F_{i_j})}(t) = \frac{\lambda_{i_j, i}}{\sum_{j=1}^n \lambda_{i_j, i}} \dot{P}_{F_i}(t)$$

the integration of which proves the theorem, if at $t = 0$ system is initialized, i.e. put into operation for the first time. ♦

Based on the above result, a computing procedure will be described in the next section.

Before going into it, we observe that, as it has been shown in [4], the classical Markov approach in terms of macrostates can be formally obtained starting from the microstates equations if the conditions of mergeability are satisfied.

These conditions imply, in particular, that different microstates pertaining to the same macrostate (and then which have to be merged into it) cannot have trajectory conditioned transition rates.

This is an intrinsic limitation of the Markov approach, which, for example, requires an unlimited set of spares for repairable systems.

3. A PROPOSED COMPUTING PROCEDURE

With reference to a system of order n , the following approximate computing procedure is based on the fundamental theorem given in the previous section.

One has immediately, see also eq. (5):

$$P_0(t) = e^{-(\sum \lambda)_0 t} \quad (8)$$

Then, of course, the cumulative probability of all the remaining microstates (sons of s_0), is given by:

$$P_{F_0}(t) = 1 - P_0(t) \quad (9)$$

Now, according to the referred theorem, the overall probability of the general subtree having root on the first level is:

$$P_{1+F_1}(t) = \frac{\lambda_{1,0}}{(\sum \lambda)_0} [1 - P_0(t)] \quad (10)$$

At this point the subtrees are examined to analyze if they can cumulatively be associated to a certain condition for the system (e.g. system failure) relevant to be problem at hand. In this case the pertinent probability computed by (10) is associated with such condition.

Among the remaining 1st level subtrees, the one having the highest probability is selected, and the probability of its root s_1 is evaluated by solving its equation of kind (6).

The probability $p_1(t)$ is therefore evaluated and associated to the condition which is satisfied in s_1 .

Furthermore $p_{F_1}(t)$ is evaluated by subtracting $p_1(t)$ from $p_{1+F_1}(t)$ of eq. (10).

At this point the 2nd level subtrees originated by the evaluated s_1 are considered by applying the same procedure.

These subtrees are examined in terms of system conditions to associate them to any condition. The remaining are considered together with those not evaluated of the 1st level, in search of the new "pivot".

The procedure is continued until when the sum of all the considered microstates differs from one by an amount lower than an allowed error.

To the purpose of a better appreciation of the above procedure, let us consider, as an example, the system in fig. 1, to which the following values are applied:

$$\lambda_A = 200 \text{ FIT}$$

$$\lambda_B = 3500 \text{ FIT}$$

$$\lambda_C = \begin{cases} 350 \text{ FIT (IF B GOOD)} \\ 3500 \text{ FIT (IF B FAILED)} \end{cases}$$

It is requested to evaluate the system reliability for $t = 10 \text{ years} = 87600 \text{ hrs.}$

The applicable dynamic tree is as in fig. 3, from which one has immediately, according to (5):

$$P_0 = 0.701328 \quad (\text{System good})$$

Again, according to (9):

$$P_{F_0} = 1 - 0.701328 = 0.298672$$

and, due to the fundamental theorem:

$$P_{1+F_1} = \frac{350}{4050} \times 0.298672 = 0.0258111$$

$$P_{2+F2} = \frac{3500}{4050} \times 0.298672 = 0.258111$$

$$P_{3+F3} = \frac{200}{4050} \times 0.298672 = 0.0147492$$

The "pivot" is the subtree originated by s_2 , then:

$$P_2(t) = 3500 \times 10^{-9} P_0(t) - 3700 \times 10^{-9} P_2(t)$$

The solution of which yields:

$$P_2 = 0.218357 \quad (\text{System good})$$

Therefore:

$$P_{F2} = 0.258111 - 0.218357 = 0.039754$$

P_{F2} is the new pivot, then one has:

$$P_{6+12} = \frac{3500}{3700} \times 0.039754 = 0.0376051 \quad (\text{System failed})$$

$$P_{7+13} = \frac{200}{3700} \times 0.039754 = 0.0021489$$

If P_F and P_G indicate respectively the evaluated probabilities of success and failure for the system, one has:

$$P_G = P_0 + P_2 = 0.919685 \quad (\text{System good})$$

$$P_F = P_{6+12} = 0.0376051 \quad (\text{System failed})$$

$$P_F + P_G = 0.9572901 \quad (\text{overall evaluated probability})$$

Therefore if the evaluation would be truncated at this point the resulting maximum error would be:

$$1 - P_F - P_G = 0.0427099$$

If the above maximum error is not acceptable by the required accuracy, as it is probably the case with a figure like the one above, one has to select the new pivot between: P_{1+F1} , P_{3+F3} and P_{7+13}

In our case it is P_{1+F1} , then one has to compute P_1 , which results to be:

$$P_1 = 0.0218357 \quad (\text{System good})$$

and:

$$P_{F1} = 0.0039754$$

Therefore:

$$P_{4+10} = \frac{3500}{3700} \times 0.0039754 = 0.0037605 \quad (\text{System failed})$$

$$P_{5+11} = \frac{200}{3700} \times 0.0039754 = 0.0002149$$

The new evaluated probabilities are:

$$P_G = 0.919685 + 0.0218357 = 0.9415207$$

$$P_F = 0.0376051 + 0.0037605 = 0.0413656$$

$$P_F + P_G = 0.9828863$$

and the maximum truncation error is:

$$0.0171137$$

Again one has to decide if the above maximum error is acceptable or not; in case it is not the pivot has to be identified among: P_{3+F3} , P_{5+11} , P_{7+13}

It is P_{3+F3} , then a new evaluation step is started, giving:

$$P_3 = 0.0123955 \quad (\text{System good})$$

$$P_{F3} = 0.0023537$$

$$P_{8+14} = \frac{350}{3850} \times P_{F3} = 0.0002139$$

$$P_{9+15} = \frac{3500}{3850} \times P_{F3} = 0.0021398 \quad (\text{System failed})$$

$$P_G = 0.9415207 + 0.0123955 = 0.9539162$$

$$P_F = 0.0413656 + 0.0021398 = 0.0435054$$

$$P_F + P_G = 0.9974216$$

maximum error = $1 - (P_F + P_G) = 0.0025784$

If the above is still not acceptable, one notice that the new pivot among:

P_{5+11} , P_{7+13} , P_{8+14} , is P_{7+13}

Then:

$$\begin{aligned}
 P_7 &= 0.0019339 && \text{(System good)} \\
 P_{13} &= 0.000215 && \text{(System failed)} \\
 P_G &= 0.9539162 + 0.0019339 = 0.9558501 \\
 P_{13} &= 0.0435054 + 0.000215 = \underline{0.0437204} \\
 P_F + P_G &= 0.9995705 \\
 \text{Maximum error} &= 0.0004295
 \end{aligned}$$

The above procedure may be continued by considering P_{5+11} and P_{8+14} .

Of course the above is only an example, which has been selected to allow hand computations.

4. CONCLUSIONS

The present paper has pointed out the peculiar properties of the microstate concept, and of the consequent *dynamic tree* structure, as alternative to the classical macrostate approach. The resulting method has to be considered as a further tool to evaluate the system reliability characteristics, it is particularly important in case of trajectory dependent behaviours. A computing procedure has been proposed, which shows that, despite the number of microstates, in general infinite, a truncated analysis can be performed to the desired degree of approximation.

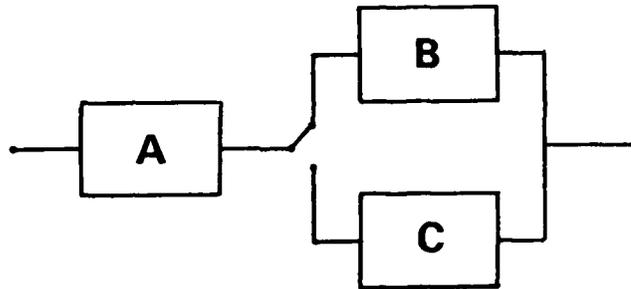
REFERENCES

1. D.E. Knuth: *Fundamental Algorithms*, Vol. I - Addison Wesley, 1968
2. V. Amoa, R. Somma: *Proprietà strutturali dell'Albero Dinamico e determinazione delle informazioni Caratteristiche del Microstato* - Selenia Spazio NT-0160/82, June 2, 1982
3. V. Amoa, R. Somma: *Considerazioni sulla Valutazione dell'Affidabilità di Sistemi di Grandi Dimensioni* - Selenia Spazio NT-0129/82 Jan. 8, 1982
4. V. Amoa, R. Somma: *Il modello di Markov tradizionale come modello ridotto dell'Albero Dinamico* - Selenia Spazio NT-0138/82, Mar. 8, 1982

THE EVOLUTIVE STATE:

A NEW DEFINITION OF SYSTEM STATE TO INCLUDE SEQUENCING OF EVENTS

LET US START WITH THE FOLLOWING EXAMPLE:



B: INITIALLY OPERATING ITEM

C: COLD STAND-BY REDUNDANCY OF B

A: SWITCHING FUNCTION. IF A FAILS IT REMAINS IN THE POSITION OCCUPIED BEFORE FAILURE

• THEREFORE: (0 = GOOD, 1 = FAILED)

A FAILS BEFORE B

| A | B | C | SYSTEM |
|---|---|---|--------|
| 1 | 1 | 0 | 0 |
| 1 | 1 | 1 | 0 |

THE OPERATING CONDITIONS OF THE COMPONENTS (I.E. THE CLASSICAL SYSTEM STATE) ARE IDENTICAL, BUT THE SYSTEM IS IN DIFFERENT OPERATING CONDITIONS.

A FAILS AFTER B

| A | B | C | SYSTEM |
|---|---|---|--------|
| 1 | 1 | 0 | 1 |
| 1 | 1 | 1 | 0 |

THE CLASSICAL RELIABILITY DESCRIPTION USES THE CONCEPT OF STATE:

STATE = OPERATING CONDITION OF THE COMPONENTS

<<>>

IS THIS SUFFICIENT TO COMPLETELY CHARACTERIZE THE SYSTEM?

TO ANSWER THIS QUESTION LET US CONSIDER THE EXAMPLE AND THE FOLLOWING SITUATION:

A AND B FAILED

IS THE SYSTEM FAILED OR NOT?

1) IF B FAILS AFTER A, THE SWITCH CANNOT ACTIVATE THE REDUNDANT BLOCK
SYSTEM IS FAILED

2) IF A FAILS AFTER B, THE SYSTEM IS GOOD OR NOT IS SO IS THE ITEM C.

IN GENERAL :

THERE ARE SITUATIONS, WHICH ARE NOT FULLY DESCRIBED BY THE OPERATING CONDI-
TIONS.

RATHER THEY REQUEST THE SEQUENCING OF EVENTS TO BE COMPLETELY EVALUATED.

↓
EVOLUTIVE STATE ≙

INFORMATION CONTENT OF THE CLASSICAL STATE (I.E. OPERATING CONDITION OF THE
COMPONENTS)

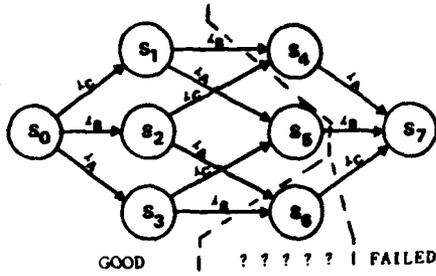
+
SEQUENCING OF EVENTS LEADING TO IT

CLASSICAL STATE ≙ MACROSTATE

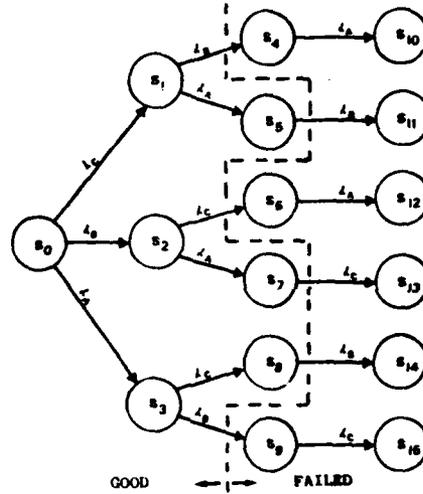
EVOLUTIVE STATE ≙ MICROSTATE

THE SET OF THE MACROSTATES IS A PARTITION OF THE SET OF MICROSTATES.

WITH REFERENCE TO THE EXAMPLE



CLASSICAL TRANSITION DIAGRAM



DYNAMIC TREE

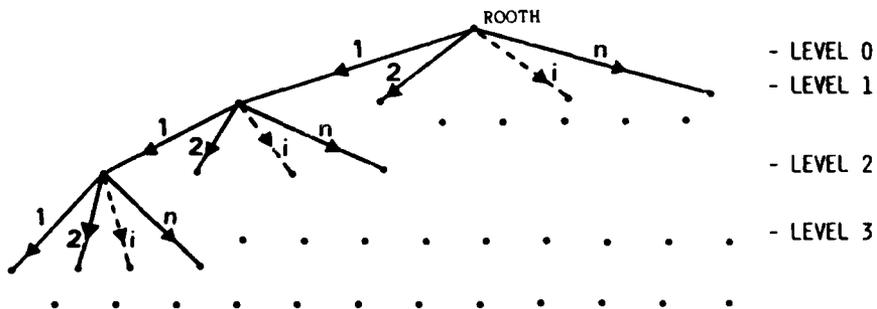
PARTITION OF THE SET OF MICROSTATES

- $S_0 \Rightarrow \{s_0\}$
- $S_1 \Rightarrow \{s_1\}$
- $S_2 \Rightarrow \{s_2\}$
- $S_3 \Rightarrow \{s_4, s_6\}$
- $S_4 \Rightarrow \{s_3\}$
- $S_5 \Rightarrow \{s_5, s_8\}$
- $S_6 \Rightarrow \{s_7, s_9\}$
- $S_7 \Rightarrow \{s_{10}, s_{11}, s_{12}, s_{13}, s_{14}, s_{15}\}$

IN GENERAL

FOR A SYSTEM OF ORDER N, THE MICROSTATE TRANSITION DIAGRAM IS AN N-ARY TREE;

NODES \Rightarrow MICROSTATES
 BRANCHES \Rightarrow TRANSITIONS



LET US CONSIDER THE GENERAL MICROSTATE S_i :

$$P_i(\tau + \Delta\tau) = [1 - (\Sigma\lambda)_i \Delta\tau] P_i(\tau) + \lambda_{iJ} P_J(\tau) \Delta\tau$$

WHERE:

$P_i(\tau)$ PROBABILITY THAT THE SYSTEM IS IN THE MICROSTATE S_i AT THE TIME τ

$(\Sigma\lambda)_i$ = CUMULATIVE TRANSITION RATE GOING OUT FROM S_i

λ_{iJ} = TRANSITION RATE FROM S_J TO S_i

$\Delta\tau$ = TIME INTERVAL

FROM THE ABOVE EQUATION:

$$\dot{P}_i(\tau) = -(\Sigma\lambda)_i P_i(\tau) + \lambda_{iJ} P_J(\tau)$$

IN PARTICULAR :

FOR S_0

$$\dot{P}_0(\tau) = -(\Sigma\lambda)_0 P_0(\tau)$$

IF (AS USUAL) $P_0(0) = 1$

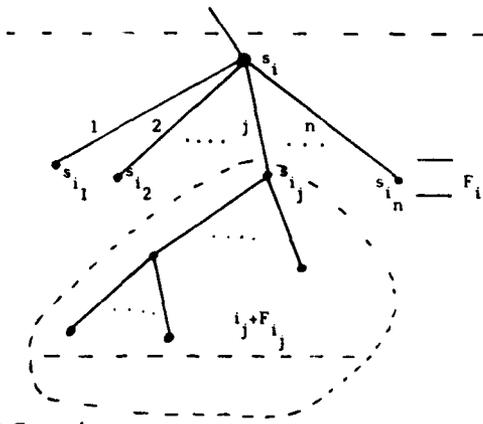
$$P_0(\tau) = e^{-(\Sigma\lambda)_0 \tau}$$

FOR S_i (MICROSTATE OF LEVEL 1):

$$\dot{P}_i(\tau) = -(\Sigma\lambda)_i P_i(\tau) + \lambda_{i,0} e^{-(\Sigma\lambda)_0 \tau}$$

FROM WHICH $P_i(\tau)$ IS DERIVED, TO BE USED TO FIND $P_{i1}(\tau)$, AND SO ON.

THEOREM (FUNDAMENTAL)



$$P_{(i_j + F_{i_j})}(\tau) = \frac{\lambda_{i_j, i}}{\sum_{j=1}^n \lambda_{i_j, i}} P_{F_i}(\tau)$$

$\lambda_A = 200$ FIT $\lambda_B = 3500$ FIT $\lambda_C = \begin{cases} 350 \text{ FIT (IF B GOOD)} \\ 3500 \text{ FIT (IF B FAILED)} \end{cases}$
 $\tau = 10$ YEARS = 87600 HRS

| | | | | | |
|---|--|--|--|---|-----|
| $P_0 = e^{-\sum \lambda \tau}$ 0.701328 (G) $-x-$ $P_{F_0} = 0.298672$ | $P_{1+F_1} = 0.0258111$ $P_{2+F_2} = 0.258111$ ▲ $P_{3+F_3} = 0.0147496$ | $P_{1+F_1} = 0.0258111$ $P_2 = 0.218357$ (G) $P_{F_2} = 0.039754$ ▲ $P_{3+F_3} = 0.0147496$ | $P_{1+F_1} = 0.0258111$ ▲ $P_{6+12} = 0.0376051$ (F) $P_{7+13} = 0.0021489$ $P_{3+F_3} = 0.0147496$ | $P_1 = 0.0218357$ (G) $P_{F_1} = 0.0039754$ $P_{7+13} = 0.0021489$ $P_{3+F_3} = 0.0147496$ | ETC |
| | $P_G = 0.701328$ $P_F = 0$ $e_T = 0.298672$ | $P_G = 0.919685$ $P_F = 0$ $e_T = 0.080315$ | $P_G = 0.919685$ $P_F = 0.0376051$ $e_T = 0.0427099$ | $P_G = 0.9415207$ $P_F = 0.0376051$ $e_T = 0.0208742$ | |

CONCLUSION

WHY MICROSTATES?

THERE ARE SYSTEM CHARACTERIZATIONS WHICH DEPEND ON SEQUENCE OF EVENTS

E.G. COST

EACH MICROSTATE CAN BE ASSIGNED

A COST $\rightarrow C_i$

A PROBABILITY AT T $\rightarrow P_i$



AVERAGE COST AT T = $\sum_i P_i C_i$

OPTIMISATION DES SPÉCIFICATIONS D'UN SYSTÈME AIR-SURFACE LASER

par

Jean-Pierre MURGUE
THOMSON-CSF Div. AVS/Dép. AVG
52, rue Guynemer
92132 - ISSY LES MOULINEAUX (FRANCE)

et

Didier EVRARD
MATRA - Secteur Militaire
37, Avenue Louis Bréguet
78140 - VELIZY (FRANCE)

Résumé

Les spécifications d'un système d'armes doivent bien cerner les besoins des Etats Majors, afin de les satisfaire dans les meilleures conditions de coût-efficacité.

Cette présentation donne un exemple des études préalables qui ont été nécessaires pour spécifier un système d'attaque Air-Surface à guidage laser.

Elle porte sur un cas opérationnel particulier très dimensionnant qui est l'attaque des vedettes lance-missiles à la mer.

Elle montre que le bombardement est une solution économique et qu'il est possible d'optimiser en coût-efficacité la définition de la nacelle de désignation laser pour le bombardement qui demande un débattement important de la ligne de visée vers l'arrière, et celle du guidage de la bombe qui doit rester à très basse altitude pour satisfaire aux conditions météorologiques envisagées, et être suffisamment précise pour atteindre des cibles aussi petites et rapides.

I - PREAMBULE

Les premières armes à guidage laser ont fait leur apparition en opération au début des années 70.

Dès 1972, au Vietnam, les excellents résultats obtenus avec les bombes guidées laser PAWEMAY démontrèrent largement la validité des études de coût-efficacité sur lesquelles s'appuyait le développement de ces armes et qui avaient mis en évidence les gains très importants que l'on pouvait obtenir avec ce type d'armement par rapport aux charges analogues non munies de guidage, même larguées de façon très précise par les calculateurs des avions modernes sur des objectifs terrestres fixes.

La confiance mise depuis dans le guidage laser par les Etats Majors et l'évolution des technologies optroniques ont conduit à repenser ce concept pour l'adapter également à l'attaque des objectifs terrestres de dimensions réduites, et aux objectifs marins mobiles.

C'est ainsi qu'a été étudié et optimisé le système d'armement guidé laser français utilisant l'illuminateur ATLIS de la Société THOMSON-CSF, l'engin AS 30 laser de la Société AEROSPATIALE et les bombes guidées laser de la Société MATRA.

Cette présentation porte, parmi les études préalables de coût-efficacité qui ont conduit aux spécifications de ce système, sur un cas opérationnel très dimensionnant intéressant la Marine : l'attaque de vedettes lance-missiles.

On verra ainsi comment ont été optimisées certaines caractéristiques de la nacelle ATLIS et la définition de la bombe MATRA, à propos de laquelle nous nous limiterons au kit de guidage. En effet, le Symposium n'étant pas classifié, nous ne pourrions pas traiter des travaux concernant les charges militaires, qui, bien évidemment, ont été conduits parallèlement par la Société MATRA pour optimiser l'arme dans son ensemble. Nous devons également renoncer à citer des informations à caractère confidentiel, ce qui nous conduira le plus souvent à présenter nos résultats sous forme qualitative.

Après avoir rappelé rapidement le concept du système d'armement guidé laser français, nous vous présenterons l'exemple de nos travaux choisis pour cet exposé, en suivant le plan ci-après :

- Description de la prise en compte :
 - . la cible et son environnement tactique,
 - . les conditions météorologiques,
 - . le scénario d'attaque,
 - . le choix du type d'arme,

- . les caractéristiques de la nacelle
 - . le choix des principes de guidage des armes.
- Enfin, nous concluons sur l'utilité de ces travaux préliminaires.

II - SYSTEME D'ARMEMENT GUIDE LASER FRANCAIS

Ce système est un système d'attaque Air-Surface pour avion monoplace, constitué par :

- la nacelle de désignation laser aéroportée ATLAS réalisée par THOMSON-CSF,
- des armes à guidage laser de deux types :
 - . Missile AS 30 Laser de l'AEROSPATIALE avec Autodirecteur Laser ARIEL de THOMSON-CSF,
 - . Bombes guidées laser BGL de MATRA avec détecteur-écartomètre laser de THOMSON-CSF.

Ce système est aujourd'hui entièrement développé et opérationnel.

La nacelle comporte un système "télévision et laser" de haute performance, dont la ligne de visée est capable d'un grand débattement ne laissant qu'un faible cône aveugle vers l'arrière. Sa télévision, qui travaille aussi bien dans la bande visuelle que dans le proche infrarouge, permet au pilote de détecter et de reconnaître sa cible à des distances nettement supérieures à celles obtenues à l'oeil nu. Il déclenche alors la poursuite télévision du point visé sur la cible. Le laser sert alors à la télémétrie pour calculer l'entrée dans le domaine de tir de l'arme qui peut être tirée aussitôt.

Dès le tir, le pilote dégage un virage serré pendant que la poursuite maintient la ligne de visée de la nacelle sur la cible. En temps utile pour le guidage final de l'arme, l'illumination laser est déclenchée automatiquement, créant sur le point visé une tache laser que l'Autodirecteur détecte, reconnaît et poursuit pour le guidage de l'arme.

On obtient ainsi une très grande précision à l'impact qui peut, grâce à la télévision, être observée par le pilote et enregistrée pour le compte-rendu de l'attaque.

Il faut noter qu'il est possible également de répartir les fonctions de désignation et celles de tir entre plusieurs avions, les uns porteurs de nacelles, les autres porteurs d'armes.

III - DESCRIPTION DE LA MISSION PRISE EN COMPTE

Parmi les missions nouvelles intéressant les Etats-Majors, nous avons donc retenu pour cette étude, le cas très dimensionnant de l'attaque de vedettes lance-missiles.

III-1. La cible et son environnement tactique.

La cible était un bâtiment de type OSA II navigant à grande vitesse, isolé ou en groupe, mais dans des dispositifs suffisamment dispersés pour que dans chaque passe une seule vedette soit à prendre en compte.

Ses caractéristiques maximales de manoeuvrabilité et de vitesse selon le vent et la mer ainsi que ses défenses étaient, bien sûr, définies de façon prospective et classifiée.

III-2. Les conditions météorologiques.

A partir des théâtres d'opérations envisagés, les conditions météorologiques susceptibles d'être rencontrées dans cette mission purent être appréciées.

CONDITIONS METEO WEATHER CONDITIONS

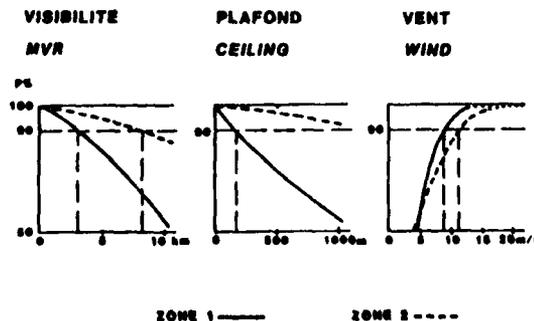


FIGURE 1

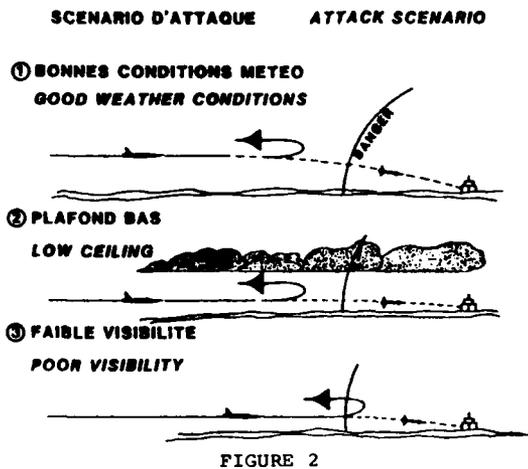
La figure 1 montre des exemples des résultats statistiques utilisés. Ils concernent la visibilité, le plafond et le vent.

Nous avons pris comme conditions nominales les limites correspondant à une probabilité de 90 %, soit, selon les zones :

- de 3 à 8,2 km pour la visibilité,
- de 180 à plus de 1000 mètres pour le plafond,
- et de l'ordre de 20 nœuds pour le vent.

III-3. Le scénario d'attaque.

Les données sur la cible et son environnement conduisaient aux conclusions suivantes sur le scénario d'attaque avec guidage laser, dans le cas principal retenu par la Marine, c'est-à-dire avec l'avion tireur assurant lui-même la désignation laser (Figure 2) :



1/ Quand les conditions météorologiques le permettent, l'attaque sera faite à l'altitude la plus favorable, à distance de sécurité.

2/ Par plafond bas, l'attaque devra être effectuée sous le plafond, depuis l'acquisition jusqu'à l'impact de l'arme, pour assurer l'illumination de la cible et l'arme devra, elle aussi, voler sous le plafond pour rester dans des conditions de propagation atmosphérique lui permettant d'acquiescer pour son guidage terminal la tâche laser sur la cible.

3/ Par mauvaise visibilité, l'avion pourra avoir à pénétrer dans le volume battu par les défenses de l'ennemi. La durée de cette pénétration devra être aussi courte que possible.

IV - LE CHOIX DU TYPE D'ARME

Ces contraintes opérationnelles permettaient de dégager deux premières spécifications pour l'arme :

- 1) Sa portée devra être suffisante pour assurer le passage à distance de sécurité.
- 2) Pour couvrir le cas du plafond bas, elle devra pouvoir être tirée et guidée à très basse altitude.

Ces spécifications étaient compatibles d'une bombe guidée capable, sans propulsion, d'un vol à très basse altitude à la portée requise, et certainement moins chère qu'un missile.

Un tel choix conduisait à traiter en terme de coût-efficacité les spécifications de la nacelle et de la bombe, en particulier les deux problèmes suivants :

- capacité de débattement de la ligne de visée de la nacelle de désignation,
- principe de guidage-pilotage de la bombe.

V - LES CARACTERISTIQUES DE LA NACELLE.

DEBATTEMENT MAXIMAL DE LA LIGNE DE VISEE DE LA NACELLE

POD LINE OF SIGHT MAXIMUM LOOK BACK ANGLE

DEBATTEMENT IMPORTANT
LOOK BACK ANGLE LARGE

PLUS FAIBLE
SMALLER

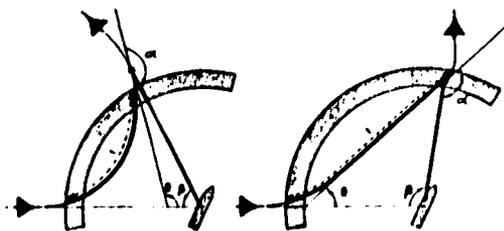


FIGURE 3

On voit sur la figure 3 que plus le débattement maximal (α) de la ligne de visée est grand, moins sera limitée l'évolution (θ) de l'avion au décollage après tir pour assurer la désignation laser jusqu'à l'impact de la bombe, et plus courte sera la durée de l'exposition (t) au feu ennemi lorsque la visibilité météorologique oblige à tirer à courte portée et à pénétrer dans le volume dangereux.

L'angle (β) de rotation de la ligne de visée entre le tir et l'impact sera aussi moins ouvert; il est en effet souhaitable que cet angle ne soit pas trop élevé pour que l'énergie laser renvoyée par la cible vers l'arme reste suffisante.

Mais l'augmentation de l'angle (α) qui, idéalement devrait être de 180°, peut coûter cher.

Il a été nécessaire d'évaluer la valeur optimale de cet angle en rapport coût/efficacité.

La figure 4 montre quelques résultats des simulations pour des bombardements à courte portée correspondant aux faibles visibilités avec pénétration dans la zone battue par les

OPTIMISATION DU DEBATTEMENT
LOOK BACK ANGLE OPTIMIZATION

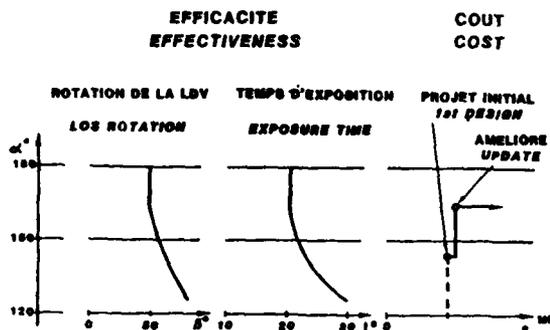


FIGURE 4

défenses ennemies. Ils montrent qu'il est particulièrement important d'augmenter l'angle α jusqu'à des valeurs proches de 180°, mais que la sensibilité des paramètres d'efficacité opérationnelle, tels que t et β , à cette augmentation s'annule pratiquement au delà de 165°.

Il nous a été facile de conclure sur ce problème quand l'étude de coût nous a montré qu'une modification de la nacelle initiale pour atteindre un angle de cet ordre conduirait à une augmentation de 10 % du coût, en apportant une diminution du temps d'exposition de l'avion de 20 %, alors qu'une modification pour un angle supérieur obligerait à changer complètement le concept, avec évidemment une augmentation très importante du coût, sans pour cela diminuer davantage le risque d'attrition.

VI - CHOIX DES PRINCIPES DE GUIDAGE DES BOMBES

VI-1. Spécifications opérationnelles des bombes guidées MATRA

Les spécifications opérationnelles des Etats Majors assignaient aux systèmes de guidage LASER MATRA deux objectifs essentiels :

- des objectifs terrestres immobiles de dimensions réduites,
- des objectifs maritimes, et plus particulièrement, des vedettes lance-missiles du type OSA II, dont la description a été faite précédemment.

Il est apparu très vite que ce dernier objectif était le plus difficile à traiter tant par ses dimensions très réduites sous certaines présentations, que par sa vitesse.

Une étude comparative en termes de coût-efficacité a été entreprise au début du développement afin de déterminer la loi de guidage la mieux appropriée, compte tenu également de la charge militaire qui représente, dans ce type d'armement, plus de 75 % de la masse totale.

Cette étude a permis de définir une solution originale, ayant un très bon rapport coût-efficacité.

VI-2. Description de la trajectoire

La trajectoire d'une arme à guidage LASER tituée de deux phases :

- une phase de préguidage, entre le point de largage et l'instant où le détecteur laser peut accrocher la cible,
- une phase autoguidée depuis l'accrochage jusqu'à l'impact.

La phase de préguidage doit pouvoir se dérouler :

- à très basse altitude lorsque le plafond nuageux l'impose,
- à plus haute altitude lorsqu'un angle d'impact final élevé est recherché et que les conditions météorologiques et l'environnement tactique de la cible le permettent.

Le guidage terminal, quant à lui, doit assurer à l'arme une précision suffisante pour atteindre la cible.

Seule l'étude relative à la composante latérale de la distance de passage sera exposée ici.

VI-3. Analyse de la précision requise

La vedette lance-missiles est un exemple de cible de dimensions relativement importantes, mais animée d'une vitesse élevée.

Elle peut se présenter à l'arme (voir Figure 5, ci-dessous) :

- par le travers,
- par l'avant ou l'arrière
- ou dans une position intermédiaire caractérisée par l'angle θ entre sa vitesse et la trajectoire de la bombe.

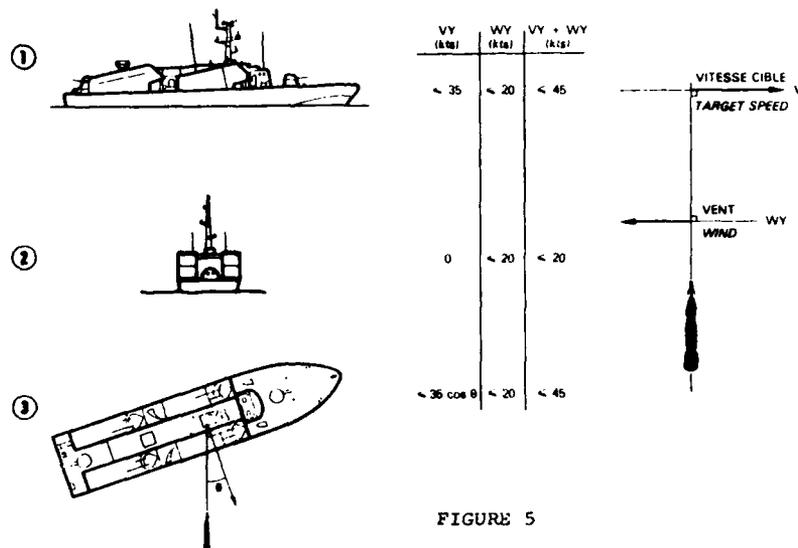


FIGURE 5

Dans le premier cas la composante perpendiculaire à la trajectoire de la bombe de la vitesse de la cible, (VY), est maximale et la vitesse du vent suivant le même axe (WY) peut s'y ajouter s'il est de face. Le vent sur le pont (VY + WY) ne peut cependant qu'exceptionnellement dépasser 45 noeuds. Mais, dans ce cas, la largeur apparente de la cible est maximale et si la tache laser est supposée au milieu du bateau, la bombe peut éventuellement faire une distance de passage non nulle tout en détruisant l'objectif grâce à sa charge militaire très efficace.

En présentation avant ou arrière, la vitesse transversale de la cible est nulle mais sa largeur apparente est minimale. La vitesse maximale du vent considérée est de 20 noeuds.

Entre ces deux présentations extrêmes, toute situation intermédiaire doit être envisagée car, pendant le vol de l'arme, le bateau a tout loisir de manoeuvrer pour éventuellement rechercher la présentation la plus favorable pour lui.

Le tracé des largeurs (figure 6) et vitesses transversales apparentes (figure 7) de la cible permet d'évaluer la distance de passage maximale pour faire impact (figure 8).

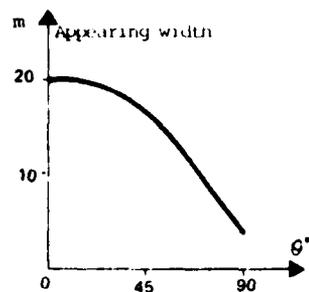


FIGURE 6

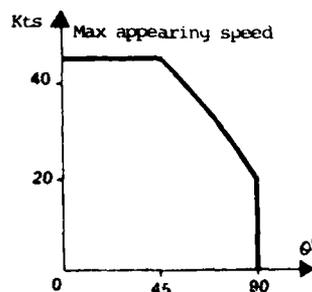


FIGURE 7

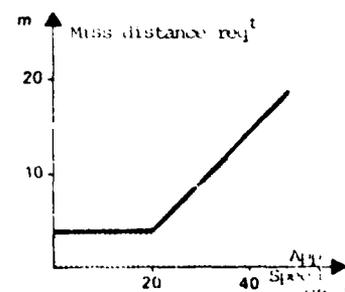


FIGURE 8

La distance de passage maximale admissible par rapport à la tache laser est donc la demi-largeur du bateau jusqu'à 20 noeuds, puis cette distance augmente.

VI-4. Définition de la loi d'auto guidage

Les objectifs de précision étant établis, il s'agissait de déterminer la loi de guidage la meilleure en termes de rapport coût/efficacité, l'efficacité étant définie comme la probabilité de faire impact, quel que soit le point du bateau.

Deux types de loi peuvent être envisagés pour cette arme :

- une loi de poursuite nécessitant un minimum d'équipements et en particulier un détecteur laser à faible coût,
- une loi de navigation proportionnelle, pure ou mixte et pondérée par des termes de poursuite, nécessitant des équipements plus complexes : autodirecteur à tête gyrocopique par exemple.

COMPARAISON DES LOIS DE GUIDAGE TERMINAL TERMINAL GUIDANCE LAWS COMPARISON

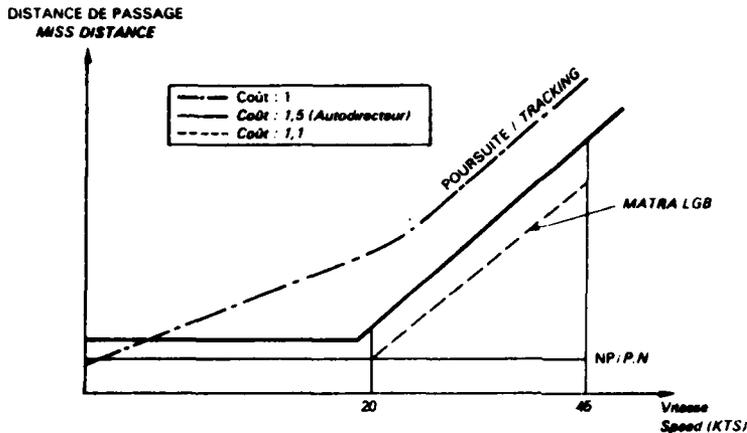


FIGURE 9

La première solution, très économique, ne répond malheureusement pas au problème car elle induit une distance de passage, proportionnelle en première approximation à la vitesse de la cible, la situation se dégradant même encore à grande vitesse à cause des saturations d'ordre aérodynamique.

La deuxième solution, utilisant un autodirecteur à tête gyrocopique amène à un surcoût de 50 % sur le système de guidage. Elle répond du point de vue précision au problème mais n'utilise pas les caractéristiques dimensionnelles de la cible et elle est donc à la limite, inutilement chère.

MATRA a donc recherché une solution intermédiaire, mieux adaptée aux spécifications de précision, qui a consisté

à améliorer de façon simple les détecteurs adaptés aux lois de poursuite afin de pouvoir compenser la fraction de la vitesse apparente de la cible correspondant au vent.

Cette amélioration est d'ailleurs également nécessaire pour les objectifs terrestres de faibles dimensions. Le surcoût entraîné par cette solution est beaucoup plus faible que pour la solution précédente : environ 10 %.

Ce rapport étant non classifié, nous ne pourrions donner que les principes de la solution adoptée.

Les détecteurs utilisés dans les bombes guidées MATRA sont :

- un écartomètre laser, monté sur girouette aérodynamique,
- un gyroscope de roulis, assurant une référence de verticale.

Le pilote est un pilote 2 axes : roulis et tangage.

La loi de guidage est une loi de poursuite avec :

- compensation de la pesanteur,
- compensation de la vitesse transverse de la cible.

Ceci est réalisé grâce au gyroscope de verticale et à l'écartomètre Laser qui a été rendu à cet effet plus performant, avec des caractéristiques de linéarité très bonnes et une grande précision, sans toutefois atteindre la complexité de l'autodirecteur. Le gyroscope de verticale, quant à lui, était déjà nécessaire pour le pré-guidage à très basse altitude.

Avec cette loi de guidage, la distance de passage reste faible pour une vitesse de cible inférieure à 20 noeuds.

Au delà, elle croît comme avec une loi de poursuite mais reste compatible de la cible envisagée.

Ainsi, en utilisant une loi d'une efficacité voisine de celle procurée par la navigation proportionnelle généralement préconisée pour traiter une cible mobile, nous avons pu réduire notablement le coût du système de guidage et donc réaliser une véritable étude d'optimisation en termes de coût/efficacité.

Les nombreux essais en vol effectués par ce système ont permis de confirmer la validité de la solution retenue et montré une précision très bonne de l'ensemble parfaitement compatible de ses objectifs opérationnels.

VII - CONCLUSION

Nous avons voulu montrer ici l'intérêt et la nécessité d'études à caractères technico-opérationnels pour aboutir à une définition cohérente et optimale en coût/efficacité d'un système complexe tel que celui de l'armement guidé LASER français, avec sa nacelle et ses armements.

Dans les cas particuliers considérés, les solutions les plus performantes où les plus complexes se sont avérées inutilement chères.

Il a été ainsi possible d'orienter la définition de la bombe vers des solutions simples mais suffisantes pour les missions assignées et d'adapter celle de la nacelle en optimisant son rapport Coût/Efficacité.

DESIGN-TO-COST (DTC) METHODOLOGY TO ACHIEVE AFFORDABLE AVIONICS

BY
ALBERT J. SHAPIRO

KEARFOTT DIVISION
THE SINGER COMPANY
150 TOTOWA ROAD
WAYNE, NEW JERSEY
07470
UNITED STATES

SUMMARY

In response to the continual exponential growth in the complexity and cost of military weapon systems, especially the electronics portions, the United States Department of Defense has implemented a "Design-to-Cost" (DTC) procurement policy. The objective of this policy is to meet ESSENTIAL and DESIRED operational requirements in the most cost effective manner by setting cost targets at the start of the procurement process.

A methodology is described for developing electronic equipment to meet DTC requirements. Specific management action is required in establishing an appropriate organization as well as procedures and guidelines for the engineering development process and subsequent production to achieve the cost targets. The critical role of computer aided design in optimizing the electronic system design is highlighted.

An example of a DTC program successfully applied to the Lightweight Doppler Navigation System (LDNS) AN/APN-128 is reviewed.

1. INTRODUCTION

The ever increasing demand for more sophisticated avionics for advanced manned and unmanned aircraft demands the use of a disciplined design and production engineering approach to assure that all essential mission requirements are being achieved for the lowest Life Cycle Cost (LCC).

U.S. Department of Defense procedures (Reference 1) for the acquisition of military equipment direct that procurement of new equipment must achieve a cost effective balance among acquisition costs, ownership costs and system effectiveness. These requirements mandate an approach to the development of new hardware which requires careful trade-offs among factors affecting performance and cost. Cost has always been a consideration, but as requirements escalate, it becomes a primary design parameter which ranks equally with performance. In recognition of the need to contain the cost of new hardware, the military customer has implemented a new acquisition technique, namely, the Design-to-Unit Production Cost (DTUPC) or Design-to-Cost (DTC) Procurement. This type of procurement is characterized by the following:

- unit production cost must become a primary design parameter
- essential performance criteria, unit acquisition and ownership costs must receive equal consideration
- realistic unit cost goals must be established that are achievable and affordable
- the best possible design to perform the mission must be obtained for the established unit cost goal.

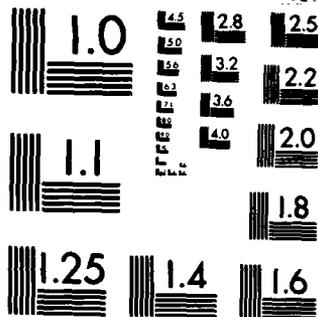
This paper describes procedures for implementing the DTC concepts that were arrived at during the development of the AN/ASN-128 Lightweight Doppler Navigation System now in production for the U.S. Army and for military forces of several other nations. It was one of several systems that was designated by the DOD as a "test bed" for evaluating the DTC concept. The techniques described in this paper should provide designers with a proven methodology to optimize avionics' cost effectiveness.

2. DESIGN-TO-COST METHODOLOGY

The sections that follow develop a suggested set of procedures, trade-offs and documentation to be used in implementing a DTC program. These are diagrammed in Figure 1 and discussed below.

2.1 SELECTION OF COST AND PERFORMANCE GOALS

The DTC process starts with the procuring agency establishing a realistic cost goal for the equipment that is to be purchased (Step 1 of Figure 1). Typically, this goal represents the amount the government is willing to pay for a unit of military equipment which meets established and measurable performance requirements at a specified production quantity and rate during a specified period of time.



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

The performance requirements are established by the technical arm of the procuring agency, and are divided into two categories: "ESSENTIAL" and "DESIRED". ESSENTIAL requirements are those elements of performance that must be met if the equipment is to be useful in the intended applications and, therefore, cannot be traded off against cost. DESIRED requirements or features can be sacrificed by the designer in order to achieve the DTC goal.

The determination of cost and performance goals by the procuring agency can be a complex and difficult trade-off study in itself. The partitioning of the available budget among the elements of a weapon system (e.g., airframe, propulsion, avionics, weapons, etc.) is a possible first step and must be based on experience and the mission requirements. This partitioning would have to be tested against estimates of the production costs of these elements. Such estimates might be based on extrapolations of the cost and performance history of similar equipments or on a survey of current suppliers to get their projections for costs and capability of "next generation" equipments. Depending on circumstances, the procuring agency may use any one or a combination of these techniques to arrive at the DTC cost goal.

The next step, the development of ESSENTIAL and DESIRED requirements, involves the cooperative participation of operational, developmental and logistics organizations within the military agency. These organizations may be required to make some difficult decisions when establishing these requirements in order to avoid the temptation of making every requirement an ESSENTIAL one. Similarly, if too many of the operational requirements are placed in the DESIRED category, there is the danger of the designer eliminating important features to meet his cost goals. Strong technical and project management control within the customer's organization must be exercised to achieve a realistic set of ESSENTIAL and DESIRED requirements.

2.2 EQUIPMENT SUPPLIERS RESPONSE TO THE DTC GOALS (Step 2 of Figure 1)

TECHNICAL - The equipment developer's activities should start with a detailed in-depth review and interpretation by program and technical management of the cost goals and performance specifications. All requirements and particularly the ESSENTIAL ones should be discussed with the customer to assure that they are clearly understood. With the help of the customer, priorities should be assigned to the DESIRED features where possible.

Some of the many factors to be considered when assigning priorities are:

- DESIRED features that may become ESSENTIAL requirements in the future when development of related architecture and/or avionics have been completed should be identified; for example, introduction of new data bus interfaces such as MIL-STD-1553.
- some seemingly expensive features may become significantly less costly by the time production starts; for example, reduction in computer memory costs as more dense LSI memories become available. The cost of RAM memory has decreased by twenty to one over the last 10 years.
- digital signal processing techniques involving microprocessors and VLSI/VHSIC technology with their vast potential for greater functional capability at lower costs will permit incorporation of features such as increased BITE and additional system functions in the near future without impacting cost or size goals.

After completing the reviews of the customer's DTC goals and requirements, they are then converted into a set of system requirements to be used by the development engineers in the detailed design phase. Later sections of this paper discuss some suggested CAE/CAD techniques to improve DTC effectiveness during the design phase.

MANAGEMENT - For DTC to be effective, a company must be ready to set up an organization that provides the functions required to properly accomplish the program goals. A full time DTC management function completely dedicated to implement and monitor the DTC concept is required in order to be successful and not merely provide lip-service to DTC. Experience has shown that a full time DTC manager, running an effective DTC program, will more than pay for that person. A typical organization chart is shown in Figure 2. The DTC manager is shown sharing the Program Management Office, thus having the required authority in directing the DTC activities, while having access to Engineering, Production, Logistics, and other program related functions and information. The DTC manager not only manages the overall DTC effort but also chairs a DTC Audit Committee that consists of the Directors of the engineering, production, logistics support and finance groups. The Audit Committee is more than just a staff function; it consists of line Directors whose personnel are assigned to the program. The Audit Committee, therefore, has the authority to react to problems by directing actions to be carried out. It is beneficial if members of the committee are directly involved in other related company programs, thus providing "cross-pollination" of ideas, techniques, and experience gained on these programs. Because the Audit Committee members represent all of the disciplines involved in developing the product they can critically evaluate the impact of trade-off decisions from many viewpoints.

It is suggested that the Audit Committee attend the System Design Review (SDR). Results of the other design reviews, such as electrical and mechanical design reviews

can be summarized for the Audit Committee for follow-on action if required. The committee should meet every month to review DTC production cost updates, and in addition should meet whenever the DTC manager requires it; e.g., when design changes have been submitted for approval prior to implementation.

In order to allow the DTC manager to exercise control over the design process, he must be provided with DTC production cost updates, which should be generated at appropriate intervals during the development program. These can be presented in a cost status and history package such as that shown in Figures 3, 4 and 5. The DTC manager must track these cost projections at sufficiently close intervals, weekly in most cases, to detect problems before designs proceed too far.

The module level data contained in Figure 3 are provided by Production Engineering using inputs from the disciplines listed on the chart, i.e., production engineers, test engineers, quality control, procurement, etc., based on the current status of the design. This Module Status chart provides a measure of how well the DTC goal will be met in terms of the difference between the target cost and current cost estimate. The column marked "Difference Between Last and Current Status" reflects the success of efforts to reduce or control costs during the program. The last three columns reflect the confidence level of the DTC estimate based on the nature of the cost data.

Figure 4 is the LRU Cost Status chart which is directly derived from the total of the module charts plus additional costs associated with the wiring, assembly and test of the LRU itself.

Finally, it is very desirable to chart the cost history of each LRU in graphic form to allow the DTC manager to monitor the cost trend relative to the target. Such a chart for a single LRU is shown in Figure 5.

FORMAL GUIDELINES

It is strongly recommended that a set of guidelines or standard procedures be developed that are documented and made available to all project personnel. These procedures can be incorporated into the company's standard policies manual, or be prepared specifically for a particular program. The standard should address overall DTC policy, assign responsibility and list specific procedures to be followed during the DTC program.

An overall description of DTC should be spelled out in a Standard Procedure Document which should begin by defining DTC and the concepts of ESSENTIAL vs. DESIRED features and the conditions under which it must be applied. It should specifically identify responsibilities for initiating the DTC program and for translating the customer's cost and performance specifications to a form suitable for the engineering groups. It must designate the approval authorities for each major element of the equipment DTC targets. The method for partitioning the system target cost among the LRU's must be defined as must the responsibilities of the cognizant Engineering and Production Departments in the total DTC process. The method to be used by the Audit Committee for reviewing cost progress must be specified as must the criteria for reiterating the design should that be necessary. The standard should be applied to all programs and projects which involve substantially new designs having significant production potential.

In addition to the project or management oriented guidelines discussed above, the design team should be provided with a set of Engineering/Manufacturing guidelines to be followed for the specific product under development. Though these guidelines may change for different products, and as technology changes, they should be established and adhered to for a specific product during its developmental phase. The guidelines are essentially sets of technical "do's" and "don'ts". Examples of "do's" are: use components from the preferred parts list, use minimum number of part types, minimize the number of separate assemblies to minimize I/O buffering and loss of speed in digital circuits. Examples of "don'ts" are: no circuit adjustments, no calibration of components or sub-assemblies, no selected parts or non-standard values for parts. Obviously, some guidelines are applicable to all or most avionic products, while others reflect a particular company's reliability, cost and logistic support experience. Once established, any deviation from the guidelines should require the approval of the DTC manager and the Audit Committee.

ENGINEERING

Based on the system requirements which were generated after consideration of the customer requirements, the overall system partitioning should be determined (Step 2 of Figure 1). System functions must be allocated to the Line Replaceable Units (LRU) and then to the modules. Partitioning should be performed to minimize hardware costs. For example, as one distributes functions among a number of LRU's the additional components and power for I/O buffering, connectors and cables (some shielded) can add cost and reduce reliability while providing no added functional capability. On the other extreme, one very large LRU seemingly attractive from an initial cost viewpoint may have severe installation problems and maintenance deficiencies and increased life cycle cost. System function partitioning, therefore, requires judgment and considerable experience with the basic hardware and its application.

The next task after function partitioning is to prepare product specifications for each LRU and each module (Step 3). These specifications should provide detailed requirements including performance, size, weight, power consumption, and inputs and outputs. Performance requirements for the modules are derived from an overall system performance analysis that in turn relates directly to the customer's system requirements. Because of the close relationship between cost and performance specifications, the latter should be generated to meet the system needs without "overkill". It is extremely desirable to make the module performance requirements directly traceable to the system requirements in order to enable rapid cost versus performance tradeoffs to be performed during LRU and module design. The module and LRU performance requirements must be approved by the DTC manager and the Audit Committee.

Once the module and LRU requirements have been determined, cost budgets can be assigned to these units. Selection of the cost numbers are based primarily on experience at this point although avionics suppliers, who have developed previous generations of the product, usually have on-going reliable research and development designs to form the basis for cost estimating. The allocation of costs to the module level is performed by the DTC manager with assistance from the designers, production and cost estimators and reviewed by the Audit Committee.

The module designers will by now have a complete set of requirements for their respective modules - cost, performance, interfaces and other characteristics - and can begin the detailed design activity (Step 4).

It is appropriate at this point to emphasize the benefits that Computer Aided Engineering (CAE) can provide not only in generating the electrical design, but also in optimizing component parts selection. A computerized component data base which contains the cost, electrical parameters and physical parameters of all components permits rapid and complete computation to evaluate the trade-offs between parameter choices and system benefits. This subject is discussed later in this paper.

The results of this design effort in the form of detailed parts lists, schematics, assembly drawings and test equipment requirements is provided to production engineering and logistics to translate into preliminary production costs and life-cycle-costs (Step 5). Performance analyses by module designers of their respective modules are reviewed by the systems group to determine the degree of compliance with overall requirements.

Step 6 of Figure 1 represents a decision point in the DTC process. The results of the design analyses are collected and reviewed in detail by the DTC manager and the Audit Committee for compliance with the performance and cost targets and to determine the next course of action. The three possible cases are diagrammed in Figure 6 and discussed below.

CASE 1 - DTC BUDGETS ARE BEING MET

This is the sought after case. Performance features and life-cycle-cost may continue to be reiterated to further optimize overall system performance and to lower production and life-cycle-cost should program schedules permit. The Program Manager would then review the overall program status with the procuring agency, and release to production when the proper point in the schedule is reached.

CASE 2 - DTC EXCEEDS BUDGET BY 5 PERCENT OR LESS

Trade-offs of cost and performance will continue until either the budget overrun is eliminated, or it is determined that further cost reduction effort is not warranted and the overrun can be absorbed by "underruns" on other modules.

CASE 3 - INDIVIDUAL DTC BUDGETS ARE EXCEEDED BY 5 PERCENT OR MORE

In those cases where an individual DTC budget exceeds its goal by 5 percent or more, the responsible performing organization will prepare a Recovery Plan and review same with the DTC Manager and the Audit Committee. Should the DTC Manager determine the total internal DTC target is not in jeopardy and the Recovery Plan is feasible on the basis of analyses and trade-offs performed, the Recovery Plan will be approved, a budget adjustment will be authorized, and the Recovery Plan will be implemented by the responsible organization.

If the DTC goal cannot be met after examination of all possibilities has been made, then the DTC Manager and the Program Manager must open a dialogue with the customer. He must be appraised of the magnitude of the problem and must be engaged in a determination of the next step. This next step could range from relief from some of the ESSENTIAL requirements to a major redirection of the program. Such a situation would arise only if some gross miscalculation of the technical difficulty of the task had been made at the outset. This is not very likely for the development of hardware using mature technology.

2.4 DTC FOLLOW-UP DURING PRODUCTION

The DTC technique for meeting tight cost and performance requirements must continue throughout the production phase.

After the design is frozen and the drawings released, the production plan, which had been generated during the development phase to support the analysis of production costs, must be updated and implemented. At this point the manufacturing cost analysis should be updated.

The detail level to which production costs must be recorded to provide control is shown in Figure 7. This is a value-added flow diagram which begins with raw material acquisition and ends with an accumulated total manufacturing cost for the product. It uses material quotations from potential vendors, in-house fabrication actuals, and test and support labor for all of the operations to be performed. Efficiency factors, assembly and test yield, and rework ratios may be derived from these data. This complete model accurately identifies the cost drivers and isolates areas for significant cost reduction opportunities. To be maximally effective, the Production Cost Tabulation of Figure 7 should be updated frequently to detect changes or trends in production costs due to component prices, producibility or yield problems.

This type of reporting must begin early in the production phase. It provides the kind of data (discrepancies, unfavorable ratios, etc.) needed to identify production problems readily.

The data provided by the Production Cost Tabulation of Figure 7 must be tracked by the production groups and reported through the use of a Producibility Analysis Report shown as Figure 8. This report may be initiated by any of the production groups such as production engineering, procurement, methods, inspection, etc., who would be identified in the "Problem Reported By" column. A proposed solution might be generated by joint action or the problem might be referred back to engineering. The "Engineering Response" would generally be in the form of a Revision Notice delineating a drawing, part or manufacturing procedure change which has received the proper DTC management reviews and approvals.

It must be pointed out that the DTC design phase results in a paper output of performance and cost estimates. It is in the production phase that the degree of DTC success is truly established as the product's actual performance is determined by ATP/Qualification Tests and/or Flight Test, and as actual costs are accrued.

3. COMPUTER-AIDED ENGINEERING TECHNIQUES

The use of computer aided engineering (CAE) has become a convenient design aid that is now used routinely throughout the electronics industry. CAE techniques are mandatory for DTC programs since they make practical the detailed statistical analyses that enable selection of the combination of components that provide the required performance at the lowest cost. CAE analyses are performed on designs before circuits are even breadboarded to minimize development cost and time. After a baseline design has been analyzed satisfactorily, breadboards are built and tested to verify the analysis.

3.1 USE OF DESK TOP CAE

Systems employing primarily analog circuitry can make use of CAE programs contained in desk-top computers. For these systems, existing computer programs such as ECAP are useful. Some systems may require special programs to be developed to perform cost versus performance trade-offs. This approach toward circuit design provides many advantages including the following:

- development is faster and costs less
- statistical analysis of component error effects are readily performed
- circuit design can be iterated until the minimum number of different types of components is achieved.
 - component tolerances can be increased by iterating the design up to the specified performance levels
 - sensitivity analyses can be performed, whereby critical components and parameters can be identified in terms of module output error
 - worst case analysis of error effects can be done to provide means for eliminating all adjustments
 - detailed design documentation can be automatically provided
 - analyses can be easily repeated if the effects of changes in module or system specifications are needed for DTC trade-offs

3.2 COMPONENT DATA BASE

Another advantage of CAE is the minimization of component types. A Preferred Parts List (PPL) generated at the outset of the program, should contain proven but state-of-the-art components, and circuit designers should be required to use these components unless proof is supplied that none of them are suitable for the specific application. For example, the PPL should include the specific resistor and capacitor

values and diode and IC types to be used to reduce the number of different active and passive elements thereby simplifying purchasing and inventory costs, and allowing low price, large quantity purchases to minimize production and life-cycle costs. A typical data base is shown in block form in Figure 9. The program should list approximately 40 parameters for each component that are printed out on three separate pages; one page lists primarily cost-related parameters, the second, performance and reliability, and the third, mechanical data such as size, weight, board area, etc. A summary sheet should be provided that includes:

- number of components by module
- the component cost by module
- the estimated area for layout
- the weight of each module
- the failure rate and MTBF by module

A copy of a typical summary sheet as used in the DTC development of the U.S. Army AN/ASN-128 is given in Figure 10. Copies of typical cost, performance and mechanical sheets are given in Figures 11, 12 and 13.

Many advantages accrue from this approach. All circuit analysis programs are automatically tied to this component data base. If a circuit design change is made, any changes in components are automatically reflected in the summary and components lists. The printout also tells the Product Engineer how many different kinds of components are being used. A Preferred Parts List is made up and circuits are redesigned, by computer, to use only those parts, if at all possible. The summary sheet tells the designer immediately his cost and performance status. Problem areas, such as excessive board area, or potential temperature problems are quickly uncovered, and corrective actions can be taken, prior to any breadboarding. Finally, a component list for parts purchasing can be made up "instantly" for the entire system since a listing of the total number of each type of part used in the system is provided as an output. Maximum advantage can, therefore, be taken of large quantity purchasing agreements.

3.3 UNIQUE CAE APPLICATIONS

Computer aided engineering is not limited to electronic circuit design and layout but can be beneficial in other areas of design. For example, microwave antenna array design has always involved some "cut and try" methods, with little flexibility in optimization because of the costly and time consuming nature of the experimental procedures. At Singer-Kearfott, a computer program has been developed which permits rapid and effective design, analysis, and fabrication of test models of a printed antenna. (Excellent correlation has been consistently obtained between predicted and measured performance.)

Figure 14 shows the inputs to this program and the results which are produced. The inputs include the physical size available for the antenna aperture, the direction of the peak of the beam and the allowable side lobe level. The program will print out the electrical performance parameters of the antenna and the physical details and dimensions of the radiating elements. The latter information is used directly to generate the artwork for etching the printed antenna and for constructing other mechanical features.

In this program, Monte Carlo techniques are used to estimate the effects of dimensional tolerances. This is preferable to the usual worst-case analysis which, when applied to microwave assemblies, has been shown to yield costly and "over-performing" designs. A flow diagram of this technique is shown in Figure 15. The key input to this program is the rms tolerance of the relevant antenna dimensions. These are applied to a random number generator to produce random sets of antenna dimensions having a variance related to the input rms tolerance. The performance range of the antennas which will be produced is then computed and the histogram shown is generated for a specified parameter, in this case the side lobe level. For the case shown, the mechanical tolerance allowed will produce a 99 percent yield of acceptable antennas. The use of conventional design techniques would incur a cost much higher than the benefit which would result from the 1 percent higher yield.

3.4 ADVANCED CAE/CAD

The above section described a CAE and data base concept successfully applied to RF and highly analog oriented systems. With electronic systems becoming more digital in nature and Printed Wiring Boards (PWB's) and Very Large Scale Integrated (VLSI) Circuits becoming ever more complex (i.e., complexity doubling every 2 years) use of a highly integrated sophisticated CAE system has become mandatory.

Typical of modern CAE tools for the design of digital electronics is the Prime 850 Computer workstation system which includes an EDMS (Electronic Design Management System) and THEMIS (The Hierarchical Event-driven Multilevel Interactive Simulator).

ELECTRONIC DATA MANAGEMENT SYSTEM (EDMS)

EDMS software integrates the electronic engineering, design, drafting, and testing information that goes from the initial electrical/electronic circuit concept to a finished proven design, providing a single Data Base Management System that can be

shared by all members of a development team.

The EDMS features two integrated data bases; an Electrical/Electronic Component Parts Library that can contain all of the standard and customized items suitable for electronic designs, and a Project Design Data Base (PDDB) that centralizes all information for a particular design, thereby permitting all users to access the same data (including cost) for consistency and accuracy. The PDDB includes the preferred parts list.

The engineer inputs the Schematic (Logic or Circuit Diagram) via a Graphic Station into the Project (Design) Data Base using libraries within the standard Controlled Component Parts Library Data Base. The Project Manager and/or DTC Manager has the freedom to assign a number of libraries or to generate an additional sub-library for new parts, if needed, for a particular project. When a Parts List is extracted from EDMS, parts are categorized as to whether they are standard or nonstandard. This Component Library Data would be controlled by Component Engineering. Users would have read-only access privileges, being restricted from doing any editing, deleting or changing of component data. The Component Library Data Base, in addition to physical characteristics and diagram presentations, can include cost related data (component cost plus assembly/test cost), reliability failure rates, and electrical characteristics.

EDMS includes several standard utility programs such as signal loading, equivalent gate summation, cross-reference checking and comparison, and power dissipation and cost summarization.

When a change is made to the project data, EDMS will automatically update all records that are impacted. The system automatically records information detailing who made a change. The entire history of a product's development and cost projections is documented and on file for project management review.

SIMULATION OF DIGITAL CIRCUITS

With the expanding use of increasingly complex LSI/VLSI digital circuits, it is no longer possible to debug a wire-wrap breadboard version of a system using custom IC's when errors may be within the custom designed IC's themselves. For designs that only use standard parts, breadboarding may not be possible if high speed logic is used. In high speed ECL logic, wire delay significantly affects circuit operation and the wire-wrap version is not an accurate model of the printed circuit version. Finally, the increase in the gate count of digital systems is making it more difficult to detect and correct errors during hardware prototyping.

The Hierarchical Event-driven Multilevel Interactive Simulator (THEMIS) permits design engineers to perform logic simulation of large complex circuits of up to 1 million gate equivalents. Models can be generated at various levels (switch, gate, functional and compiler language) enabling modeling when the actual circuit configuration is unknown or has not yet been designed.

THEMIS permits the modeling of strengths of a signal as well as its value. Additionally, the timing characteristics of models can be more thoroughly defined since there are different delays for rising and falling edges, delays associated with capacitive loading as well as charge decay often encountered in MOS circuits. Set up and hold conditions can also be described and race conditions can be detected and reported. Temperature and loading effects simulating various application environments can be added to the program. This type of CAE facility is an excellent tool for parametric analysis and DTC optimization.

Finally, the THEMIS Fault Simulator utilizes the concurrent fault simulation technique that permits the simulation of many faults in parallel. This will result in considerably faster operation as compared to existing serial fault simulators which take many CPU hours to simulate a moderately complex circuit. This capability has an important benefit for production testability at either the device or card level.

The Prime EDMS/THEMIS CAE/CAD concept described above and shown in Figure 16 provides the design engineer with the tools needed to translate their ideas into hardware in minimum time and with minimum chance for errors, while giving management the tools to control and monitor the design process, product cost and projected reliability.

4. A DESIGN-TO-COST CASE HISTORY - DEVELOPMENT OF THE AN/ASN-128 LIGHTWEIGHT DOPPLER NAVIGATION SYSTEM (LDNS)

4.1 THE AN/ASN-128

The LDNS is the result of a conscientiously applied Design-to-Cost program in which system design features were continuously traded off against cost.

To better understand the ASN-128 DTC program and its results, a brief description of the system will be given. The AN/ASN-128 Lightweight Doppler Navigation System (LDNS) is the standard airborne Doppler navigation system for the U.S. Army and for the military forces of several other countries. The LDNS consists of three units: a Receiver Transmitter-Antenna (RTA), a Signal Data Converter (SDC) and a Computer-

Display Unit (CDU). Figure 17 shows these three units.

The RTA and SDC constitute the Doppler Radar Velocity Sensor (DRVS) which continuously measures the velocity of the aircraft. The CDU provides the flight crew with controls and displays, and contains the navigation computer. With inputs from the vehicle's heading and vertical references, the ASN-128 LDNS provides accurate velocity, present position, and steering information from ground level to over 10,000 feet. It is completely self-contained and requires no ground based aids.

The entire system is solid-state, weighs less than 13.61 Kg (30 pounds), and requires less than 100 watts of 28VDC Power. Operational mean-time-between-failure is well over 1,000 hours.

Built-in-Test-Equipment isolates failures to the LRU and also to the module level to simplify maintenance and eliminate the need for field test equipment.

4.2 DEVELOPMENT HISTORY

The AN/ASN-128 was developed using the Design-to-Cost methodology described in the foregoing section. The customer, in this case the U.S. Army, provided a DTC cost target predicated on a specified production quantity and rate over a specified future time period. A performance specification delineated the ESSENTIAL and DESIRED features. (See Table I.)

These specifications were generated by the Army after a survey had been made by an Army committee assigned to review preliminary approaches offered by candidate vendors. This committee visited each vendor and performed an in-depth review of his equipment, in some cases down to the parts list level, to obtain a data base for generating the DTC specification. In this program, Singer-Kearfott was one of two suppliers selected to develop the LDNS.

Singer-Kearfott's development effort followed the DTC methodology outlined previously. Some of the design ground rules that were established early in the program are worth setting down despite their obvious nature. These were:

- Simplicity of Design - start with a simple design and keep it simple.
- State-of-the-Art Innovations - use only if they have been proven and are cost effective, not just because they are new.
- Use existing designs wherever practical - saves cost of design, test equipment, handbooks, drawings and productionizing.
- Incorporate producibility, maintainability and reliability features into the equipment design at the start

During the detailed system and hardware design phase a number of significant trade-off decisions were made that affected the cost of the LDNS. Some examples of these are:

- Automatic terrain bias compensation, which is highly preferable operationally, was achieved by using antenna beam shaping methods as the lowest cost approach compared to beam lobing and a land-sea switch
- Use of dense electronic packaging in the CDU to allow inclusion of the computer, thereby eliminating a separate computer LRU.
- Use of a single card CPU, namely the SKC-3020 which was already developed, in production and directly applicable
- BITE was incorporated to the module level to eliminate special test equipment thus reducing LCC cost
- Use of DIPs in the remainder of the system to reduce component and assembly costs
- A rechargeable battery and charging circuit to sustain the memory during power-off conditions had been considered. This was rejected in favor of a throw-away battery which was already in Army inventory, resulting in significantly reduced cost due to elimination of the charger.

In addition to the above trade-offs, the use of computer aided design and the early involvement of production experts were major contributors to meeting the cost targets. The CAE permitted minimization of mechanical and electrical tolerances and elimination of the need for trimming or adjusting on the production line.

The early involvement of production experts resulted in card designs suitable for automatic insertion of components and use with automatic test equipment.

The low parts count, and the reduction in fabrication-induced failures by automated production techniques, resulted in a predicted MTBF of over 2,000 hours without the

use of costly HiRel components. The AN/ASN-128 production cost for the initial buy of equipments was nearly 25 percent below the DTC goal originally assigned to it, and the goal itself was significantly lower than the actual cost of any Dopplers in production at that time.

4.3 LIFE CYCLE COST CONSIDERATION

The DTC program which minimized the production cost of the LDNS also recognized the requirement to minimize Cost of Ownership or Life-Cycle Costs. The Life-Cycle-Costs were defined as the production cost plus ten years of operating and maintenance costs. These costs were computed and traded off against various maintenance policies to develop the most cost effective maintenance plan.

The significant design characteristics which led to a minimum Life-Cycle-Cost are the following:

- Low Acquisition Cost
- High Reliability (MTBF)
- BITE capability to isolate failures to the module level without the use of special test equipment
- Minimum use of non-standard parts
- Low mean time to repair (MTTR)
- No trim pots to become misadjusted

4.4 MEETING THE ESSENTIAL TECHNICAL REQUIREMENTS

The ESSENTIAL and DESIRED technical requirements are listed in Table 1 together with the corresponding LDNS capabilities. It shows that the LDNS exceeded the ESSENTIAL requirements without compromising the DTC target.

The many trade-offs of features versus cost performed during the DTC effort did not affect acceptance by the user community of the LDNS as evidenced by the successful testing and placement of orders for over 3,000 systems by the U.S. Army and military agencies of eight other countries.

TABLE 1. ESSENTIAL TECHNICAL REQUIREMENTS

| <u>REQUIREMENTS</u> | <u>SPECIFICATION (ESSENTIAL/DESIRED)</u> | <u>LDNS CHARACTERISTIC</u> |
|------------------------------------|--|---------------------------------------|
| TOTAL WEIGHT | 50 LB/35 LB | 28 LB |
| SIZE | 3500 IN. ³ /2000 IN. ³ | 1191 IN. ³ |
| CDU | HEIGHT - 6 INCHES DEPTH - 8 INCHES | HEIGHT - 6 INCHES DEPTH - 8 INCHES |
| PRESENT POSITION | 2% CEP/1% CEP | 1.3% CEP |
| ALTITUDE | 2-10,000 FT/2-15,000 FT | 0-14,000 FEET |
| DESTINATIONS | 6/10 | 10 |
| PREDICTED RELIABILITY (MTBF) | 1000 HOURS | 2121 HOURS |

5. CONCLUSION

The writer is convinced that a properly implemented DTC program is a powerful technique for reducing the production and Life-Cycle-Cost of new avionics.

This paper has provided a detailed implementation methodology based upon experience gained on actual development programs. The prospective implementor of this type of program is cautioned that Design-to-Cost is not "just another" cost reduction technique, but it is one that requires strong discipline and close cooperation between the procuring agency and the contractor to keep cost as a major design requirement.

Experience has shown that several elements of DTC methodology are particularly important.

- The user must be realistic in dividing design features into ESSENTIAL and DESIRED categories - it takes a strong effort to avoid making all features ESSENTIAL
- An open two-way dialogue must occur between developer and buyer regarding the relative importance of the DESIRED features
- The user/buyer must be willing to accept the significant responsibility in selecting the right combination of ESSENTIAL and DESIRED characteristics, and is thereby exposed to the possibility of future criticism if operational experience shows that this selection was not correct.
- The developer is also faced with difficult decisions in determining which DESIRED features will be eliminated to meet the cost goal. During an evaluation of competing systems that meet the DTC goals, the difference in DESIRED features could decide the winner.

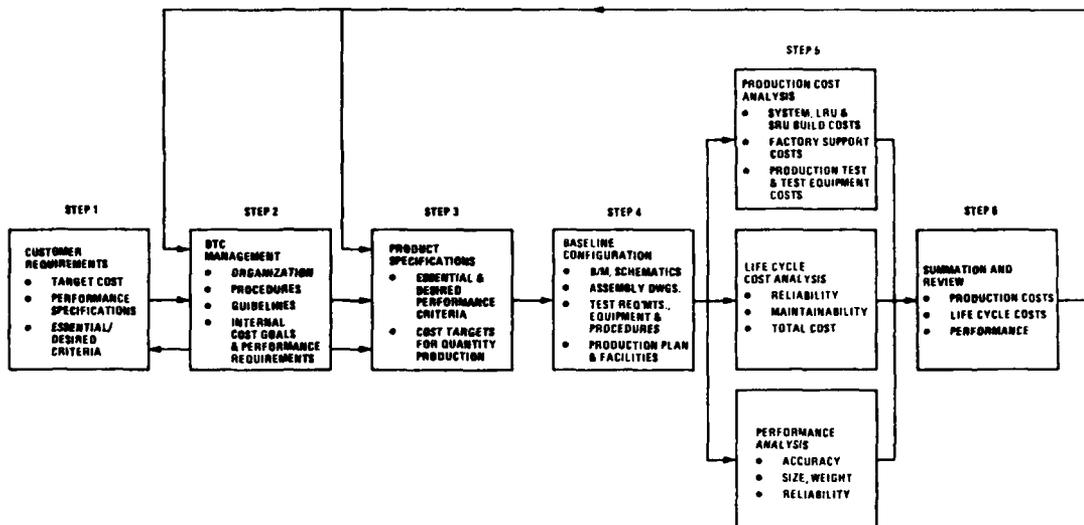
Conceived by some as a negative feature of DTC is the exposure of details of procurement and pricing policies of the customer and the vendor, respectively. The vendor witnesses first-hand the difference in opinion that can occur within the user community regarding the importance of various design features. The customer obtains details of the vendor's cost breakdown structure and pricing policies that are normally not made available and that could possibly affect the vendor's financial relationship on other contracts. Some customers and/or vendors may decide that the advantages of DTC do not justify these "exposures".

Surely the customer could avoid the above mentioned issues imposed by DTC by simply reverting to the "classical" buyer/vendor relationship. The buyer canvasses the user community with his "wish lists" and converts these into a firm system specification for the supplier to bid against. Although this procedure may seem easier for the customer and the vendor, it undoubtedly raises the cost of the final product over that of a system design resulting from a thorough trade-off between design features and their cost.

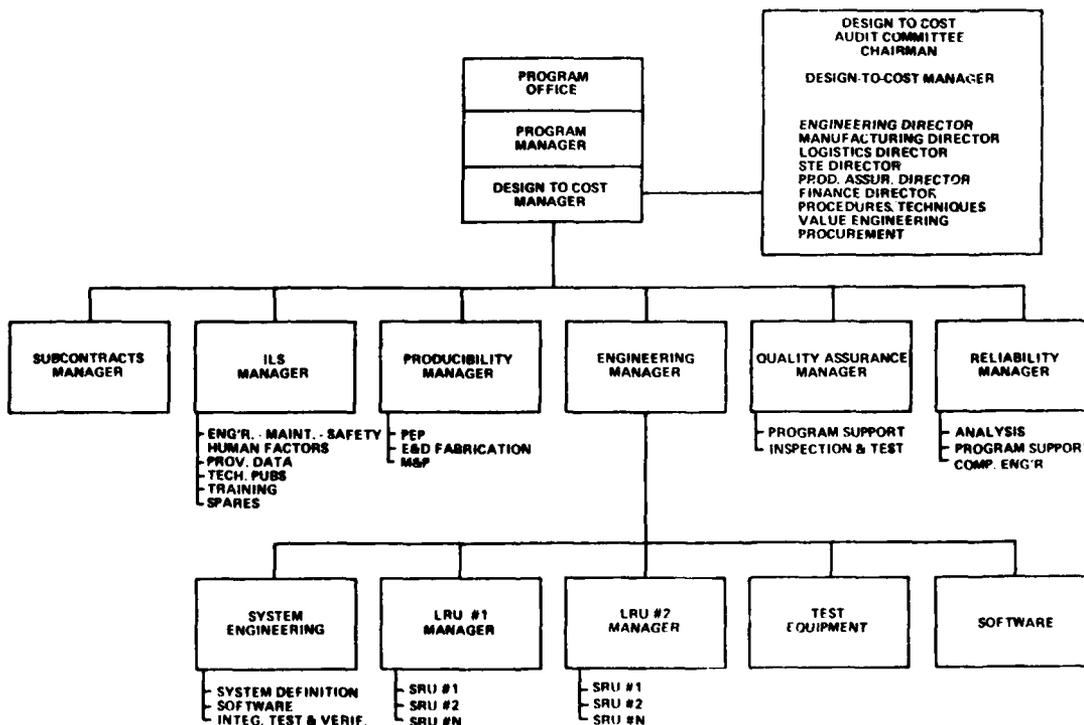
Hopefully, the techniques and experiences discussed in this paper will convince procuring agencies, supported by equipment developers, to apply the DTC concept in order to obtain cost effective and affordable avionic systems.

REFERENCE:

1. Department of Defense Directive, "Major System Acquisitions", Number 5000.1, March 29, 1982.



SYSTEM DESIGN APPROACH BLOCK DIAGRAM
 FIGURE 1



DTC ORGANIZATION AND AUDIT COMMITTEE
 FIGURE 2

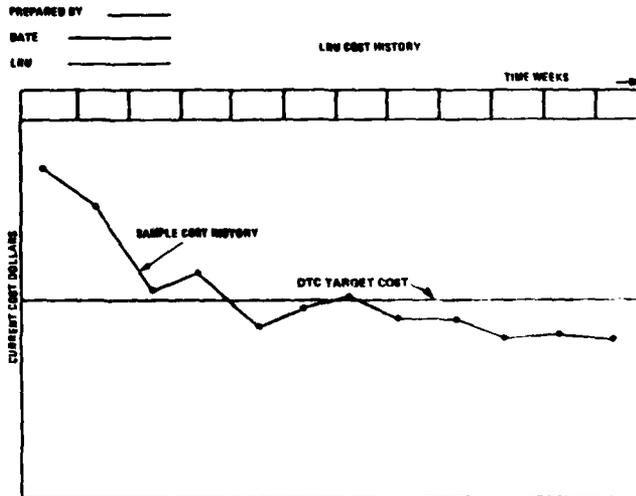
MODULE COST STATUS

| MODULE PREPARED BY _____ DATE _____ | TARGET COST | DIFF. BETWEEN TARGET & CURRENT STATUS | CURRENT STATUS DATE _____ | DIFF. BETWEEN LAST & CURRENT STATUS | CURRENT COST DATA | | |
|---|-------------|---------------------------------------|---------------------------|-------------------------------------|-------------------|--------|--------|
| | | | | | ESTIMATED | QUOTED | ACTUAL |
| 1. PURCHASED PARTS | | | | | | | |
| 2. ASSEMBLY | | | | | | | |
| 3. FABRICATED PARTS | | | | | | | |
| 4. INSPECTION | | | | | | | |
| 5. TEST | | | | | | | |
| 6. FACTORY SUPPORT | | | | | | | |
| 7. LABOR BURDEN | | | | | | | |
| TOTAL | | | | | | | |

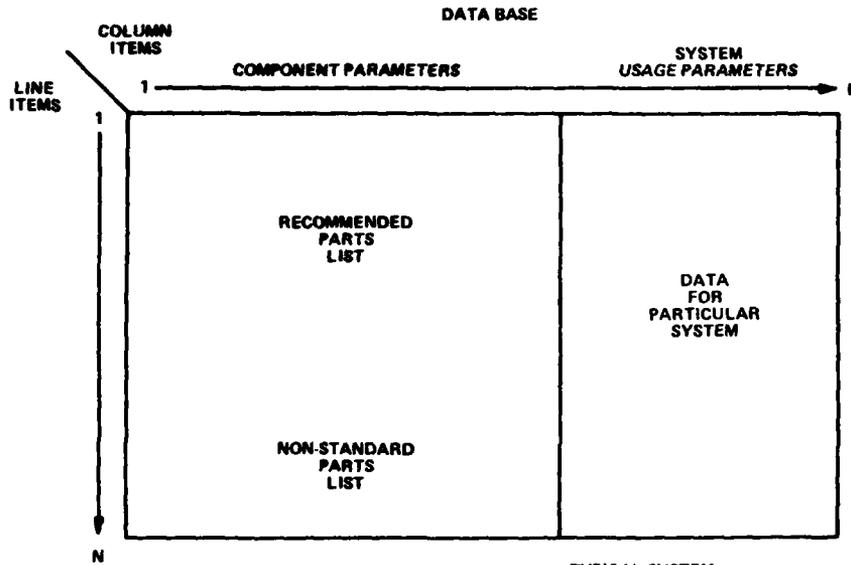
**MODULE COST STATUS REPORT
FIGURE 3**

| PREPARED BY _____ DATE _____ | LRU NO. 1 | | | | LRU NO. 2 | | | |
|---------------------------------|-------------|---------------------------------------|---------------------|-------------------------------------|-------------|---------------------------------------|---------------------|-------------------------------------|
| | TARGET COST | DIFF. BETWEEN TARGET & CURRENT STATUS | CURRENT STATUS DATE | DIFF. BETWEEN LAST & CURRENT STATUS | TARGET COST | DIFF. BETWEEN TARGET & CURRENT STATUS | CURRENT STATUS DATE | DIFF. BETWEEN LAST & CURRENT STATUS |
| 1. PURCHASED PARTS | | | | | | | | |
| 2. ASSEMBLY | | | | | | | | |
| 3. FABRICATED PARTS | | | | | | | | |
| 4. INSPECTION | | | | | | | | |
| 5. TEST | | | | | | | | |
| 6. FACTORY SUPPORT | | | | | | | | |
| 7. LABOR BURDEN | | | | | | | | |
| TOTAL | | | | | | | | |

**LRU COST STATUS REPORT
FIGURE 4**



**COST HISTORY REPORT
FIGURE 5**



TYPICAL COMPONENT PARAMETERS

- DESCRIPTION
- VALUE
- TOLERANCE
- VENDOR
- COST
- LEAD TIME
- MIL SPEC
- FAILURE RATE
- LENGTH
- WIDTH
- HEIGHT
- AREA
- WEIGHT
- LABOR HOURS

TYPICAL SYSTEM USAGE PARAMETERS

- QUANTITY USED/MODULE
- TOTAL QUANTITY
- TOTAL FAILURE RATE
- TOTAL COST

COMPONENT DATA BASE
FIGURE 9

8/31/72 10142

| TOTAL NUMBER OF COMPONENTS: | | | | | | | |
|---|----------|----------|--------------|-----------------|---------------|--------------|----------|
| TRACKER | POST-IF | PRE-IF | POWER SUPPLY | TIMER INTERFACE | SWITCH DRIVER | SYSTEM TOTAL | |
| 235 | 109 | 76 | 130 | 65 | 68 | 827 | |
| COST (DOLLARS) : | | | | | | | |
| TRACKER | POST-IF | PRE-IF | POWER SUPPLY | IFACE | SWITCH DRIVER | SYSTEM TOTAL | |
| 40,682 | 45,777 | | | 18,630 | 34,582 | 9,106 | 223,589 |
| HEIGHT OF CON. | | | | | | | |
| TRACKER | POST-IF | PRE-IF | POWER SUPPLY | TIMER INTERFACE | SWITCH DRIVER | SYSTEM TOTAL | |
| .642 | .60 | .327 | 1.600 | .263 | .399 | .123 | 3.964 |
| FAILURE RATE (FAILURES/MILLION HOURS) : | | | | | | | |
| TRACKER | POST-IF | PRE-IF | POWER SUPPLY | TIMER INTERFACE | SWITCH DRIVER | SYSTEM TOTAL | |
| 65.3615 | 81.7431 | 67.5368 | 115.0978 | 8.4868 | 12.6485 | 94.2776 | 394.8585 |
| MHP (HOURS) : | | | | | | | |
| TRACKER | POST-IF | PRE-IF | POWER SUPPLY | TIMER INTERFACE | SWITCH DRIVER | SYSTEM TOTAL | |
| 18099.04 | 10386.87 | 21035.49 | 8778.12 | 117036.04 | 74079.83 | 10687.00 | 4577.70 |

SUMMARY
FIGURE 10

| COST INFORMATION | | | | VENDOR #1 | VENDOR #1 PART NUMBER | QUANTITY | COST (VENDOR #1) TIMES # OF COMPONENTS USED IN SYSTEM | | | | | | | |
|------------------|-------|--------|---------|-----------|-----------------------|----------|---|---------|-----|----------|----------|------|------------|------|
| RECORD | COMP | DESCR | TYPE | GENERIC | VALUE | UNITS | TOL | MILSPEC | REV | MIL. NO. | QTY. NO. | STAT | FAIL. RATE | TYPE |
| 2 | AMP | OP | GEN | MC15300 | - | | | | | | | | | |
| 4 | AMP | OP | OPN | 101 | - | | | | | | | | | |
| 5 | AMP | OP | GEN | 101A | - | | | | | | | | | |
| 6 | AMP | OP | MIL/PER | 100 | - | | | | | | | | | |
| 9 | CAP | PTM/RJ | | | .000 | UP | | | | | | | | |
| 10 | CAP | PTM/RJ | | | .01 | UP | | | | | | | | |
| 12 | CAP | PIVCO | CFM | C400 | 10-125 | PP | | | | | | | | |
| 14 | CAP | PIVCO | CFM | C400 | 1200-NE5 | PP | | | | | | | | |
| 22 | CAP | PIVCO | NICA | C405 | 1-10 | PP | | | | | | | | |
| 23 | CAP | PIVCO | NICA | C405 | 1-300 | PP | | | | | | | | |
| 24 | CAP | PIVCO | NICA | C405 | 20-02 | PP | | | | | | | | |
| 25 | CAP | PIVCO | NICA | C405 | 91-300 | PP | | | | | | | | |
| 30 | CAP | PIVCO | NICA | C406 | 030-0700 | PP | | | | | | | | |
| 31 | CAP | PIVCO | NICA | C406 | 030-0700 | PP | | | | | | | | |
| 32 | CAP | PIVCO | POLY | | .015 | UP | | | | | | | | |
| 34 | CAP | PIVCO | POLY | | .01 | UP | | | | | | | | |
| 36 | CAP | PIVCO | POLY | | | UP | | | | | | | | |
| 37 | CAP | PIVCO | POLY | | | UP | | | | | | | | |
| 38 | CAP | PIVCO | TANT | | | UP | | | | | | | | |
| 40 | CAP | PIVCO | TANT | | | UP | | | | | | | | |
| 42 | CAP | PIVCO | TANT | | | UP | | | | | | | | |
| 43 | CAP | PIVCO | TANT | | | UP | | | | | | | | |
| 44 | CAP | PIVCO | TANT | | | UP | | | | | | | | |
| 46 | CAP | PIVCO | TANT | | | UP | | | | | | | | |
| 49 | CAP | PIVCO | TANT | | | UP | | | | | | | | |
| 50 | CONTR | BRIT | | 111 | | | | | | | | | | |
| 52 | CONTR | BRIT | | 54193 | | | | | | | | | | |
| 53 | CONTR | BRIT | | 93L10 | | | | | | | | | | |
| 55 | CONTR | BRIT | | MC310L | | | | | | | | | | |
| 57 | CONTR | DECADE | | MC316 | | | | | | | | | | |
| 58 | CONTR | PHASE | | MC308 | | | | | | | | | | |
| 59 | CONTR | PHASE | | 11751A | | | | | | | | | | |
| 61 | CONTR | CONTR | | L73070 | | | | | | | | | | |
| 64 | CONTR | RECT | | 111106 | | | | | | | | | | |
| 67 | CONTR | RECT | | 11751A | | | | | | | | | | |
| 68 | CONTR | RECT | | 11751A | | | | | | | | | | |
| 70 | CONTR | RECT | | 11751A | | | | | | | | | | |
| 73 | CONTR | RECT | | 11751A | | | | | | | | | | |
| 75 | CONTR | RECT | | 11751A | | | | | | | | | | |
| 77 | CONTR | RECT | | 11751A | | | | | | | | | | |
| 80 | CONTR | RECT | | 11751A | | | | | | | | | | |
| 84 | CONTR | RECT | | 11751A | | | | | | | | | | |
| 87 | CONTR | RECT | | 11751A | | | | | | | | | | |
| 89 | CONTR | RECT | | 11751A | | | | | | | | | | |
| 90 | CONTR | RECT | | 11751A | | | | | | | | | | |

COST INFORMATION
FIGURE 11

| SPECIFICATIONS | | | | | | | | | | | | | | |
|----------------|-------|--------|---------|---------|----------|-------|-----|---------|-----|----------|----------|------|------------|------|
| RECORD | COMP | DESCR | TYPE | GENERIC | VALUE | UNITS | TOL | MILSPEC | REV | MIL. NO. | QTY. NO. | STAT | FAIL. RATE | TYPE |
| 2 | AMP | OP | GEN | MC15300 | - | | | | | | | | | |
| 4 | AMP | OP | OPN | 101 | - | | | | | | | | | |
| 5 | AMP | OP | GEN | 101A | - | | | | | | | | | |
| 6 | AMP | OP | MIL/PER | 100 | - | | | | | | | | | |
| 9 | CAP | PTM/RJ | | | .000 | UP | | | | | | | | |
| 10 | CAP | PTM/RJ | | | .01 | UP | | | | | | | | |
| 12 | CAP | PIVCO | CFM | C400 | 10-125 | PP | | | | | | | | |
| 14 | CAP | PIVCO | CFM | C400 | 1200-NE5 | PP | | | | | | | | |
| 22 | CAP | PIVCO | NICA | C405 | 1-10 | PP | | | | | | | | |
| 23 | CAP | PIVCO | NICA | C405 | 1-300 | PP | | | | | | | | |
| 24 | CAP | PIVCO | NICA | C405 | 20-02 | PP | | | | | | | | |
| 25 | CAP | PIVCO | NICA | C405 | 91-300 | PP | | | | | | | | |
| 30 | CAP | PIVCO | NICA | C406 | 030-0700 | PP | | | | | | | | |
| 31 | CAP | PIVCO | NICA | C406 | 030-0700 | PP | | | | | | | | |
| 32 | CAP | PIVCO | POLY | | .015 | UP | | | | | | | | |
| 34 | CAP | PIVCO | POLY | | .01 | UP | | | | | | | | |
| 36 | CAP | PIVCO | POLY | | | UP | | | | | | | | |
| 37 | CAP | PIVCO | POLY | | | UP | | | | | | | | |
| 38 | CAP | PIVCO | TANT | | | UP | | | | | | | | |
| 40 | CAP | PIVCO | TANT | | | UP | | | | | | | | |
| 42 | CAP | PIVCO | TANT | | | UP | | | | | | | | |
| 43 | CAP | PIVCO | TANT | | | UP | | | | | | | | |
| 44 | CAP | PIVCO | TANT | | | UP | | | | | | | | |
| 46 | CAP | PIVCO | TANT | | | UP | | | | | | | | |
| 49 | CAP | PIVCO | TANT | | | UP | | | | | | | | |
| 50 | CONTR | BRIT | | 111 | | | | | | | | | | |
| 52 | CONTR | BRIT | | 54193 | | | | | | | | | | |
| 53 | CONTR | BRIT | | 93L10 | | | | | | | | | | |
| 55 | CONTR | BRIT | | MC310L | | | | | | | | | | |
| 57 | CONTR | DECADE | | MC316 | | | | | | | | | | |
| 58 | CONTR | PHASE | | MC308 | | | | | | | | | | |
| 59 | CONTR | PHASE | | 11751A | | | | | | | | | | |
| 61 | CONTR | CONTR | | L73070 | | | | | | | | | | |
| 64 | CONTR | RECT | | 111106 | | | | | | | | | | |
| 67 | CONTR | RECT | | 11751A | | | | | | | | | | |
| 68 | CONTR | RECT | | 11751A | | | | | | | | | | |
| 70 | CONTR | RECT | | 11751A | | | | | | | | | | |
| 73 | CONTR | RECT | | 11751A | | | | | | | | | | |
| 75 | CONTR | RECT | | 11751A | | | | | | | | | | |
| 77 | CONTR | RECT | | 11751A | | | | | | | | | | |
| 80 | CONTR | RECT | | 11751A | | | | | | | | | | |
| 84 | CONTR | RECT | | 11751A | | | | | | | | | | |
| 87 | CONTR | RECT | | 11751A | | | | | | | | | | |
| 89 | CONTR | RECT | | 11751A | | | | | | | | | | |
| 90 | CONTR | RECT | | 11751A | | | | | | | | | | |

COMPONENTS INFORMATION
FIGURE 12

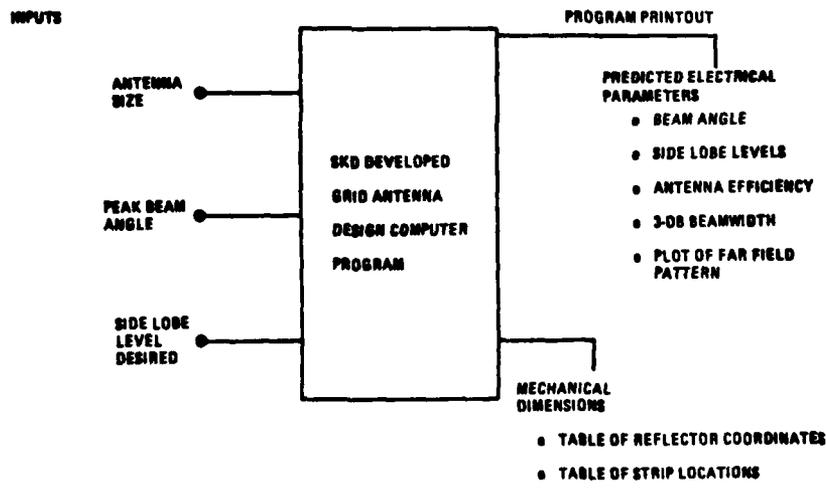
Reproduced from
best available copy.

NUMBER OF COMPONENTS PER MODULE

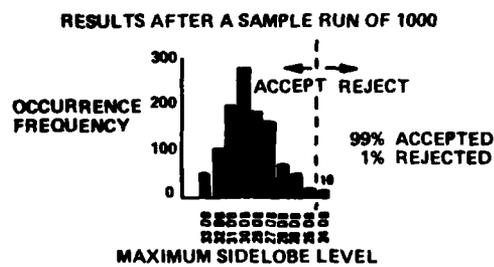
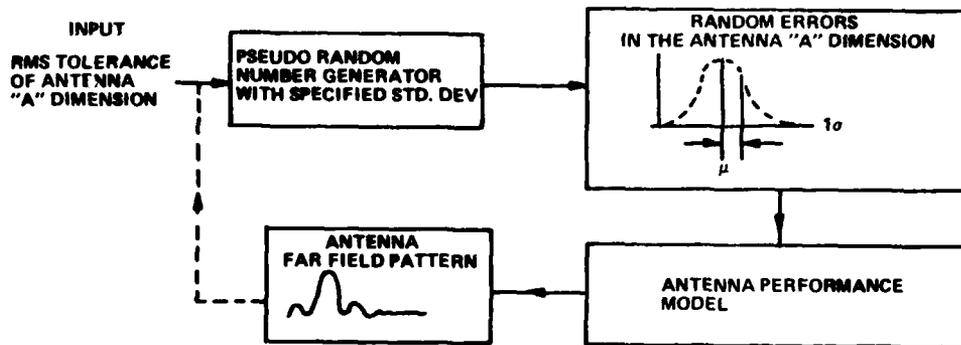
| MODULE COMP | DESCR | TYPE | GENERIC | VALUE | LEN | WID | HT | AREA | UT | TMR | POIF | PRIF | P.SUP | TMR | DATA | BUF | CAL | S4844 | TOTAL |
|-------------|--------|--------|---------|----------|------|------|----|------|-------|------|------|------|-------|-----|------|-----|-----|-------|-------|
| 8 AMP | SP | GEN | HE18900 | | 7099 | | | .235 | .0023 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 16 |
| 8 AMP | SP | GEN | 101 | | 7099 | | | .235 | .0023 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 2 |
| 8 AMP | SP | GEN | 101A | | 7099 | | | .235 | .0023 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 6 |
| 8 AMP | SP | HT/PER | 100 | | 7099 | | | .235 | .0023 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| 9 CAP | POTVAR | | | .000 | .400 | .300 | | .43 | .109 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 3 |
| 10 CAP | POTVAR | | | .1 | 1.00 | | | .43 | .109 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 12 CAP | FIXED | CEM | CM85 | 10-175 | .2 | .1 | | .000 | .000 | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 3 |
| 14 CAP | FIXED | CF-M | CM86 | 1200-825 | .3 | .1 | | .13 | .04 | 27 | 12 | 11 | 21 | 0 | 0 | 0 | 0 | 3 | 162 |
| 22 CAP | FIXED | NICA | CM85 | 1-10 | .400 | .17 | | .43 | .109 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| 23 CAP | FIXED | NICA | CM85 | 1-300 | .400 | .17 | | .43 | .109 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| 24 CAP | FIXED | NICA | CM85 | 20-82 | .400 | | | .43 | .109 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 |
| 25 CAP | FIXED | NICA | CM85 | 91-300 | .400 | | | .43 | .109 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 9 |
| 30 CAP | FIXED | NICA | CM85 | 430-477 | .400 | | | .43 | .109 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 21 |
| 31 CAP | FIXED | NICA | CM85 | 430-477 | .400 | | | .43 | .109 | 1 | 13 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 10 |
| 33 CAP | FIXED | POLY | | .01 | | | | .000 | .000 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 34 CAP | FIXED | POLY | | .01 | | | | .000 | .000 | 5 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 |
| 34 CAP | FIXED | POLY | | .02 | | | | .000 | .000 | 5 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 |
| 37 CAP | FIXED | POLY | | 1.0 | | | | .000 | .000 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 13 |
| 39 CAP | FIXED | TANT | 3500 | | 1 | | | .151 | .159 | .021 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 2 |
| 40 CAP | FIXED | TANT | 3500 | | 1 | | | .151 | .159 | .021 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 2 |
| 41 CAP | FIXED | TANT | 3400 | | 1 | | | .151 | .159 | .021 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 2 |
| 42 CAP | FIXED | TANT | 3500 | | 1 | | | .151 | .159 | .021 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 2 |
| 43 CAP | FIXED | TANT | 3400 | | 1 | | | .151 | .159 | .021 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 2 |
| 44 CAP | FIXED | TANT | 3400 | | 1 | | | .151 | .159 | .021 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 2 |
| 46 CAP | FIXED | TANT | 3400 | | 1 | | | .151 | .159 | .021 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 2 |
| 49 CAP | FIXED | TANT | 3400 | | 1 | | | .151 | .159 | .021 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 2 |
| 50 COMP | AM | | | | 7099 | .370 | | .105 | .0023 | | | | | | | | | | |
| 52 COMP | AM | | | | DL16 | | | .6 | .0052 | | | | | | | | | | |
| 53 COMP | AM | | | | DL16 | | | .6 | .0052 | | | | | | | | | | |
| 55 COMP | AM | | | | DL16 | | | .6 | .0052 | | | | | | | | | | |
| 57 COMP | AM | | | | DL16 | | | .6 | .0052 | | | | | | | | | | |
| 58 LET | AM | | | | DL16 | | | .6 | .0052 | | | | | | | | | | |
| 59 DIODE | RECT | JAN | | | DL16 | | | .6 | .0052 | | | | | | | | | | |
| 61 DIODE | RECT | JAN | | | DL16 | | | .6 | .0052 | | | | | | | | | | |
| 64 DIODE | RECT | JAN | | | DL16 | | | .6 | .0052 | | | | | | | | | | |
| 67 DIODE | RECT | JAN | | | DL16 | | | .6 | .0052 | | | | | | | | | | |
| 69 DIODE | RECT | JAN | | | DL16 | | | .6 | .0052 | | | | | | | | | | |
| 69 DIODE | RECT | JAN | | | DL16 | | | .6 | .0052 | | | | | | | | | | |
| 70 DIODE | RECT | JAN | | | DL16 | | | .6 | .0052 | | | | | | | | | | |
| 73 DRIVER | TRF | JAN | | | DL16 | | | .6 | .0052 | | | | | | | | | | |
| 75 DRIVER | TRF | JAN | | | DL16 | | | .6 | .0052 | | | | | | | | | | |
| 77 PET | SW | | | | DL16 | | | .6 | .0052 | | | | | | | | | | |
| 82 PET | JUNCT | N | | | DL16 | | | .6 | .0052 | | | | | | | | | | |
| 83 PET | JUNCT | N | | | DL16 | | | .6 | .0052 | | | | | | | | | | |
| 84 PET | JUNCT | N | | | DL16 | | | .6 | .0052 | | | | | | | | | | |
| 86 PET | JUNCT | N | | | DL16 | | | .6 | .0052 | | | | | | | | | | |
| 87 PET | JUNCT | N | | | DL16 | | | .6 | .0052 | | | | | | | | | | |
| 88 PET | JUNCT | N | | | DL16 | | | .6 | .0052 | | | | | | | | | | |
| 89 FILTER | SW | | | | DL16 | | | .6 | .0052 | | | | | | | | | | |
| 90 FILTER | SW | | | | DL16 | | | .6 | .0052 | | | | | | | | | | |
| 92 PULLER | SW | | | | DL16 | | | .6 | .0052 | | | | | | | | | | |

- TKR = TRACKER
- POIF = POST-IF
- PRIF = PRE-IF
- P.SUP = POWER SUPPLY
- TMR = TIMER
- DATA = DATA ACCUMULATOR
- BUF = BUFFER
- CAL = CAL. CONSTANTS
- SWDRY = SWITCH DRIVER

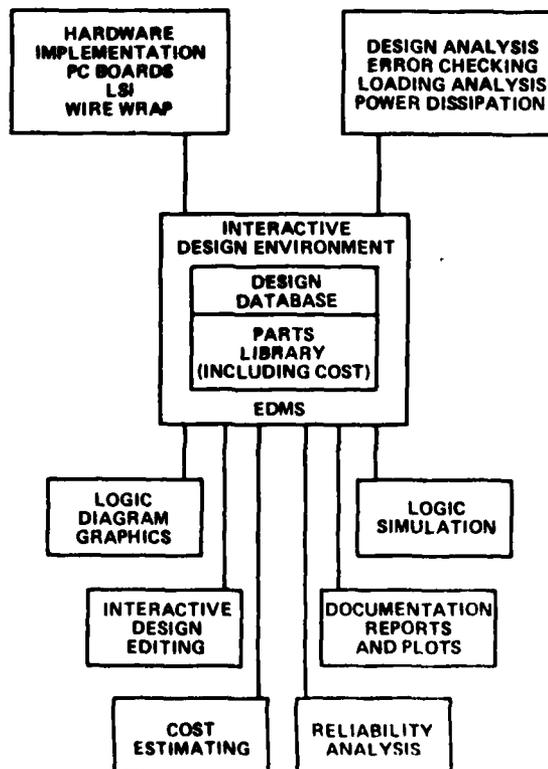
MECHANICAL INFORMATION
FIGURE 13



KEARFOTT RADIATING GRID DESIGN PROGRAM
FIGURE 14



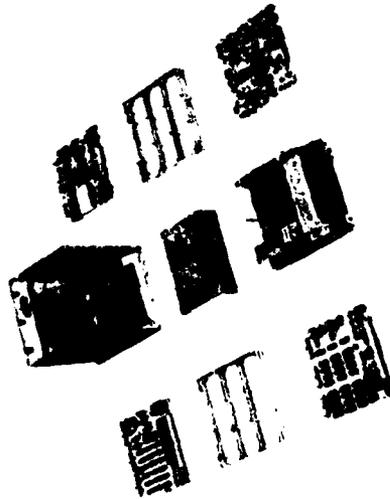
MONTE CARLO ANALYSIS
FIGURE 15



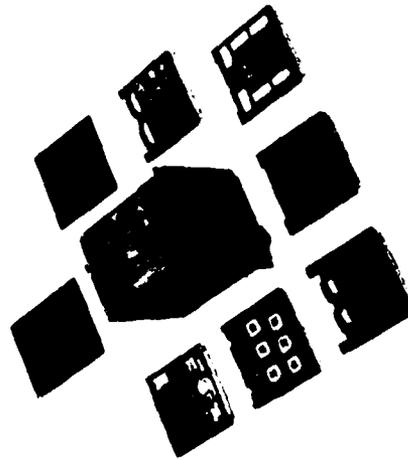
PRIME - EDMS/THEMIS
FIGURE 16

SINGER
KEARFOTT DIVISION

**THE AN/ASN-128
LIGHTWEIGHT DOPPLER NAVIGATION SYSTEM**



**SIGNAL DATA
CONVERTER**



COMPUTER-DISPLAY



**RECEIVER
TRANSMITTER
ANTENNA**

FIGURE 17

**PREGUIDAGE ET PILOTAGE
A L'AIDE DE SENSEURS INERTIELS COMPOSITES**

par

Mr J. Resseguier
SAGEM
BP 51
F.95612 Cergy Pontoise CEDEX
France

RESUME

Pour résoudre le problème du préguidage de petits engins équipés d'autodirecteur, SAGEM a conçu et fabriqué SUR MESURE des équipements SIGAL (Système Intégré Gyro-Accélérométrique Lié), compatibles avec les contraintes de volume, de maintenabilité et de prix.

Ces systèmes utilisent :

- des senseurs inertiels "composites" à la fois gyromètre et accéléromètre,
- un bouclage multiplexé,
- une organisation liée (STRAP DOWN),
- une technologie électronique adaptée.

1 - OBJECTIFS (Figure 2)

1.1 - Le développement de l'Inertie permet d'équiper des véhicules très divers.

Il y a une vingtaine d'années, l'objectif était l'obtention de performances compatibles avec les missions des engins balistiques et avions de combat.

1.2 - Ensuite, le progrès technologique entraînant une réduction des volumes et des coûts permet de doter d'un système à inertie des véhicules moins sophistiqués.

Le système SAGEM MSD, construit autour du gyromètre GSD (présenté en 1980 à une conférence AGARD), a démontré, au cours d'essais officiels au Centre d'Essais en Vol, sa capacité :

- d'alignement autonome en gyrocompas,
- de navigation en inertie pure,

avec une précision globale meilleure que 3 nautiques/heure.

1.3 - Dans le domaine des performances moyennes, pour obtenir un meilleur compromis coût-performance, SAGEM a ensuite proposé des équipements utilisant le GSL dont la classe est de l'ordre de quelques degrés par heure. Ses applications sont très nombreuses (références d'attitude, stabilisation...).

1.4 - Aujourd'hui, l'extension vers les petits engins, pour résoudre le problème du préguidage, donne de plus en plus d'importance aux contraintes (figure 3) :

- prix,
- volume,
- grande dynamique,
- très faible temps de réaction,
- simplicité-maintenabilité.

Les quantités à fabriquer sont plus importantes et les performances moins critiques.

Ces considérations obligent à concevoir des équipements en :

- limitant le nombre des composants (électromécaniques et électroniques),

- éliminant les réglages coûteux au profit de compensations par ordinateur,
- rendant automatiques les fabrications et les contrôles,
- s'adaptant au volume disponible à bord du missile.

L'objet de ce document est de décrire une réalisation correspondant à ce besoin.

2 - SENSEUR COMPOSITE - PRINCIPE

2.1 - Définition

Par "senseur composite" on entend un capteur inertiel qui, à la fois gyromètre et accéléromètre, est bien adapté à un volume réduit et à un prix faible du système de préguidage.

Plusieurs principes sont connus. Dans l'exemple décrit on ajoute la fonction accéléromètre à un gyromètre en maîtrisant les effets des accélérations linéaires (balourd).

A l'aide de 3 senseurs composites 2 axes, on réalise un système de mesure triaxial.

2.2 - Principe du système de mesure triaxial

Nous supposons connu le principe du gyroscope à suspension dynamique accordée, utilisé en mode asservi (mesure gyrométrique).

Lorsque sur un tel gyro (figure 4) on décale suivant la direction de l'axe de rotation le centre de masse G du volant (rotor) par rapport au centre de suspension O, l'appareil devient sensible non seulement aux vitesses angulaires autour de ses deux axes de mesure, mais également aux accélérations linéaires le long de ces mêmes axes.

Une vitesse angulaire Ω_x et une accélération Γ_x engendrent un couple C d'axe OY tel que :

$$\overline{C}_y = \overline{H} \cdot \overline{\Omega}_x + m \cdot \overline{OG} \cdot \overline{\Gamma}_x$$

(H = moment cinétique du volant)

(m = masse du volant)

Les tensions mesurées u_x et u_y sont des fonctions composites de Ω et Γ :

$$u_x = A_x \Omega_x + B_x \Gamma_x$$

$$u_y = A_y \Omega_y + B_y \Gamma_y$$

A ce stade, il est impossible de calculer Ω_x , Γ_x , Ω_y , Γ_y .

Cette difficulté disparaît avec 3 capteurs (figure 5) dont les axes de spin sont parallèles aux axes OX, OY et OZ et dont les coefficients A et B sont judicieusement choisis.

Aux erreurs de mesure près, les 3 capteurs délivrent les tensions suivantes :

Capteur G1 :

$$u_{1y} = A_{1y} \Omega_y + B_{1y} \Gamma_y$$

$$u_{1z} = A_{1z} \Omega_z + B_{1z} \Gamma_z$$

Capteur G2 :

$$u_{2z} = A_{2z} \Omega_z + B_{2z} \Gamma_z$$

$$u_{2x} = A_{2x} \Omega_x + B_{2x} \Gamma_x$$

Capteur G3 :

$$u_{3x} = A_{3x} \Omega_x + B_{3x} \Gamma_x$$

$$u_{3y} = A_{3y} \Omega_y + B_{3y} \Gamma_y$$

On peut calculer $\Omega_x, \Omega_y, \Gamma_x, \Gamma_y$ et Γ_z si les déterminants :

$$A_i B_j - A_j B_i$$

ne sont pas nuls.

Cette condition est satisfaite avec la configuration :

- capteur G1 : gyroscope non balourdé,
- capteurs G2 et G3 : gyroscopes balourrés d'une valeur identique, mais de moments cinétiques opposés (inversion du sens de rotation).

$$\text{Soit : } A_{1y} \approx A_{1z} \approx A$$

$$B_{1y} \approx B_{1z} \approx 0$$

$$A_{2z} \approx A_{2x} \approx -A_{3x} \approx -A_{3y} \approx A$$

$$B_{2z} \approx B_{2x} \approx B_{3x} \approx B_{3y} \approx B$$

Les coefficients A et B sont optimisés en fonction du profil de vol du missile afin de caractériser le rapport des sensibilités aux accélérations linéaires et aux vitesses angulaires.

Par exemple : $1 \text{ g} \longleftrightarrow 10^\circ/\text{s}$.

2.3 - Senseur composite GSL82

Le composant de base est un gyroscope accordé 2 axes miniaturisé ; son schéma est classique.

A partir de ce composant, SAGEM a étudié et fabriqué un senseur composite (gyroscope balourrés), le GSL 82.

La sensibilité accélérométrique est réalisée par décalage entre le centre de masse de l'élément sensible (toupie) et le centre de l'articulation (joint flexible).

Les ordres de grandeur des caractéristiques physiques sont indiqués sur la figure 6 :

- l'encombrement est faible,
- le rendement du moteur couple de précession est excellent,
- la vitesse de rotation élevée permet d'obtenir facilement la bande passante nécessaire,
- le couple du moteur élimine tout risque de désynchronisation,
- un boîtier électronique incorporé transforme le signal du détecteur en une tension continue dont le facteur d'échelle est normalisé.

3 - BOUCLAGE DU SENSEUR COMPOSITE GSL 82

3.1 - Objectifs

Le bouclage des capteurs obéit aux règles suivantes :

- commande du moteur couple en analogique. Cette méthode, la plus simple, reste la plus satisfaisante compte tenu des durées des vols et des tendances à osciller des gyroscopes deux axes,
- digitalisation des informations (vitesse angulaire et accélération linéaire) dans la boucle, pour assurer l'égalité des intégrales des signaux analogiques et numériques, sous une forme compatible avec une liaison numérique bus,
- multiplexage des informations des 3 senseurs,
- filtrage, réjection, assurés par le calculateur.

Par ailleurs :

- l'équipement est miniaturisé,
- la cadence des informations destinées au pilotage est élevée (de l'ordre de 500 Hz).

3.2 - Schéma de bouclage

Le bouclage est schématiquement représenté figure 7.

3.3 - Asservissement des 6 chaînes

On se reportera au schéma synoptique, figure 8. Trois capteurs GSL 82 sont asservis par la chaîne multiplexée :

- un multiplexeur 16 voies, commandé par le calculateur, explore en séquence :
 - . les 6 tensions continues représentant les attitudes des éléments sensibles des 3 capteurs,
 - . les 4 tensions continues représentant les températures internes des capteurs et la température ambiante,
 - . le faux zéro de la chaîne électronique de décodage,
 - . les différentes tensions d'alimentation,
 - . (la fréquence de multiplexage est voisine de 1 100 Hz),
- un convertisseur analogique-numérique digitalise les tensions ci-dessus en mots de 12 bits,
- un calculateur traite les signaux de détection des capteurs (intégration, filtrage, réseau correcteur) et calcule les commandes des moteurs couple de précession de ces capteurs (mots de 16 bits),
- un convertisseur numérique analogique de 16 bits transforme cette information en une tension analogique.

Cette tension est aiguillée vers le moteur couple correspondant au détecteur scruté, par un échantillonneur-bloqueur de précision commandé par le calculateur. Cet échantillonneur bloqueur maintient pendant une période de multiplexage le courant d'asservissement à un niveau constant.

Le signal traversant l'échantillonneur-bloqueur est amplifié par un amplificateur proportionnel ; le courant correspondant traverse le moteur couple de précession et une résistance de contre-réaction ou de mesure,

- deux niveaux de contre-réaction sont sélectionnés par le calculateur, ce qui permet un changement d'échelle automatique et confère ainsi une bonne précision aux signaux de faible niveau (pendant la majeure partie du temps de la mission).

REMARQUE : La précision de la chaîne (conformité entre le courant traversant le moteur couple de précession et la valeur correspondante digitalisée acquise par le calculateur) est donnée par :

- le décodeur,
- l'échantillonneur bloqueur,
- l'amplificateur d'asservissement.

Les autres éléments de la chaîne n'interviennent pas.

Le zéro de la chaîne, testé à chaque séquence du multiplexage, est mémorisé dans le calculateur et corrigé.

4 - TRAITEMENT INTERNE

4.1 - Fonctions assurées par le calculateur (figure 9)

Le calculateur :

- traite les signaux de détection et élabore les ordres de précession (asservissement),
- effectue les corrections thermiques et les modélisations statiques des capteurs,
- calcule les vitesses angulaires p , q , r et les accélérations linéaires a_x , a_y et a_z ,
- gère les commandes de multiplexage et de séquençement, le démarrage des toupies, la commutation de gain,
- gère l'interface,
- contrôle le bon fonctionnement,
- envoie les 3 sorties digitalisées p , q , r sur 3 convertisseurs numériques analogiques fournissant ces données sous forme analogique.

4.2 - Compensations des erreurs systématiques

4.2.1 - Les erreurs systématiques de la mesure u_x sont :

- C_x : dérive constante,
- $\alpha \Omega_y + \alpha' \Omega_z$: décalage des axes,
- $B \Gamma_y + B' \Gamma_z$: sensibilité aux composantes d'accélération linéaire sur les axes transverses.

Les valeurs des paramètres sont, pour chacun des 6 axes, mesurées lors de la calibration du bloc senseur et stockées dans la mémoire REPR0M.

4.2.2 - Le modèle peut être complété par un terme tenant compte de la variation du moment cinétique en fonction de la vitesse d'entraînement du gyro autour de l'axe de spin.

Mais il s'agit, en général, d'une perturbation de courte durée dont l'effet intégré est négligeable.

4.3 - Compensation des facteurs d'échelle

4.3.1 - Auto-échauffement

On corrige les variations de facteur d'échelle dues à l'échauffement de l'aimant du moteur-couple lorsque l'intensité d'asservissement est élevée. La durée des sollicitations est à prendre en considération.

4.3.2 - Température ambiante

La régulation thermique du bloc senseur n'est pas nécessaire. On corrige les facteurs d'échelle en fonction de la température initiale.

Une compensation dynamique, en vol, est réalisable. Elle n'est pas, en général, envisagée compte tenu :

- de la brièveté des fortes perturbations,
- de la longueur des constantes de temps thermiques,
- de la faible durée du vol.

4.4 - Compensation du faux zéro de la chaîne

A chaque séquence (soit environ à 1 000 Hz), le zéro de la chaîne est testé. Le résultat mémorisé est introduit comme un terme correctif.

4.5 - Elaboration des informations de sortie

Après correction des tensions mesurées, l'unité de traitement :

- sépare les informations :
 - . vitesses angulaires,
 - . accélérations linéaires,
- gère leur transfert par un coupleur au bus 1 553B par exemple.

Les sorties sont disponibles 500 fois par seconde environ.

5 - DESCRIPTION DU SIGNAL

(Système Intégré Gyro-Accélérométrique Lié)

5.1 - Généralités

Disposant d'un senseur (miniaturisé et ayant un grand domaine de mesure) il a été possible de concevoir un système :

- de faible volume,
- de forme adaptable à l'espace disponible,

en utilisant les techniques des systèmes liés (strap down) et en optimisant l'électronique et la connectique.

9-6

5.2 - Forme adoptée

La photographie d'un des exemples de réalisation de SIGAL (figure 10) montre à quel point les contraintes de forme ont pu être prises en compte.

Le volume de cette réalisation est de 1,5 dm³.

5.3 - Bloc senseur

En fonction des contraintes opérationnelles (vol typique du missile) l'architecture du bloc senseur peut partiellement évoluer :

- orientation préférentielle des senseurs par rapport aux axes de vitesse angulaire ou d'accélération linéaire maximale.

5.4 - Carte électronique (figure 11)

Les fonctions d'asservissement et de traitement sont regroupées sur une seule carte multicouche, constituée de quatre éléments reliés par des circuits souples.

Cette technologie, flexible-rigide, garantit un encombrement minimum et une excellente tenue en ambiance dynamique sévère.

Un seul connecteur assure l'interface du système et du missile.

Le découpage fonctionnel fait l'objet de la figure 12.

Les choix technologiques (modules hybrides - circuits prédiffusés) sont des compromis entre les contraintes de prix et de volume.

6 - SIGAL - PERFORMANCES

6.1 - Adaptabilité (figure 13)

Les performances sont partiellement adaptables aux caractéristiques du véhicule ; elles résultent de divers compromis :

- domaine de mesure en accélération linéaire et en vitesse angulaire,
- simultanéité des valeurs à mesurer ; les valeurs maximales correspondent en général à des axes orthogonaux,
- axe (roulis par exemple) des perturbations les plus importantes ou les plus critiques.

En fonction des données opérationnelles, on optimise le balourd (accéléromètre) en fonction du moment cinétique (gyromètre). Les limites sont dues aux couplages parasites.

6.2 - Ordres de grandeur

Différents systèmes SIGAL sont en développement ; les ordres de grandeur des performances sont indiqués sur la figure 14.

Les performances dépendent, entre autres, du mode d'utilisation adopté pour le système d'arme :

- dormant,
- en veille permanente.

Le temps de réaction inférieur à deux secondes est compatible avec un missile dormant.

7 - SIGAL - DEVELOPPEMENT (figure 15)

Les systèmes SIGAL sont actuellement des prototypes existant en plusieurs versions pour diverses utilisations :

Sont acquis à ce jour :

- le contrôle des performances des senseurs : environ trente GSR2 ont été testés,
- les essais statiques et dynamiques en laboratoire. Ils ont permis de développer les méthodes et les logiciels de calibration et d'évaluation,
- la qualification dynamique en laboratoire par simulation d'un profil de vol.

Les premiers essais en vol auront lieu en 1985.

8 - SIGAL - CONCLUSION (figure 16)

Les systèmes SIGAL répondent au besoin nouveau et très large du préguidage des petits engins AIR-AIR, SOL-AIR et AIR-SOL.

La disponibilité d'un capteur bien adapté à l'ensemble des spécifications permet de réaliser des systèmes "SUR MESURE" en ce qui concerne :

- la forme, l'encombrement,
- les modes d'utilisation.



Figure 1 - MIDCOURSE GUIDANCE AND CONTROL WITH COMPOSITE INERTIAL SENSORS

- 1 - GOAL
- 2 - COMPOSITE INERTIAL SENSOR (PRINCIPLE)
- 3 - CAGING LOOP
- 4 - PROCESSING UNIT
- 5 - SIGAL - DESCRIPTION
- 6 - SIGAL - PERFORMANCES
- 7 - SIGAL - DEVELOPMENT
- 8 - SIGAL - CONCLUSION



Figure 2 - SAGEM INERTIAL SENSORS

HIGH ACCURACY (0.01°/h)

- 1960 . FLOATING GYROS
PENDULOUS ACCELEROS
 - 1975 . TUNED GYROS
ELECTRICALLY SUSPENDED GYROS
- } GIMBAL PLATFORM

MEDIUM ACCURACY (0.05°/h)

- 1965 . FLOATING GYROS → GIMBAL PLATFORM
- 1975 . TUNED GYROS
 - GSP → GIMBAL PLATFORM
 - GSD → STRAP DOWN

(FLIGHT TEST OF MSD : 3 NAUT/HOUR - GYRO COMPASS - PURE INERTIAL SYSTEM)

LOW COST (1°/h ↔ 10°/h)

- 1980 . TUNED GYROS
GSL → STRAP DOWN
- STABILIZATION
- COMPOSITE SENSOR
GSL 82 → STRAP DOWN



Figure 3 - NEW GOAL

- MIDCOURSE GUIDANCE FOR SMALL MISSILES
- CONSTRAINTS
 - COST
 - VOLUME : 1.5 dm³
 - REACTION TIME : 2 sec.
 - DYNAMIC RANGE
 - SIMPLICITY
- HOW ?
 - RUGGED SENSORS
 - MINIMUM NUMBER OF INERTIAL SENSORS
 - NO EXPENSIVE ADJUSTMENTS
 - ERRORS COMPENSATION (COMPUTER)
 - SEMI-CUSTOM ELECTRONICS
 - CUSTOM DESIGNED SHAPE

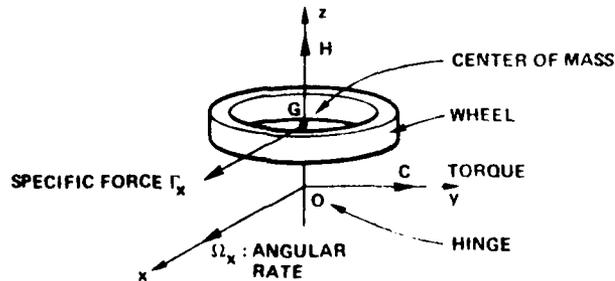


Figure 4 - COMPOSITE INERTIAL SENSOR
RATE GYRO AND ACCELEROMETER

→ 3 COMPONENTS VS 6

GYRO USING THE MASS UNBALANCE EFFECT

$$\left\{ \begin{array}{l} C = \text{MEASURED TORQUE} \\ C = \vec{H}_x \Omega_x + m \cdot \vec{OG} \times \vec{\Gamma}_x \end{array} \right.$$



ONE COMPONENT →

$$\left\{ \begin{array}{l} 2 \text{ EQUATIONS} \\ 4 \text{ UNKNOWN PARAMETERS} \end{array} \right.$$



Figure 5 - 3 AXIS COMPOSITE SENSOR

- **3 COMPONENTS** →
- 6 EQUATIONS
- 6 UNKNOWN PARAMETERS
($\Omega_x, \Omega_y, \Omega_z, \Gamma_x, \Gamma_y, \Gamma_z$)

● **MECHANIZATION**

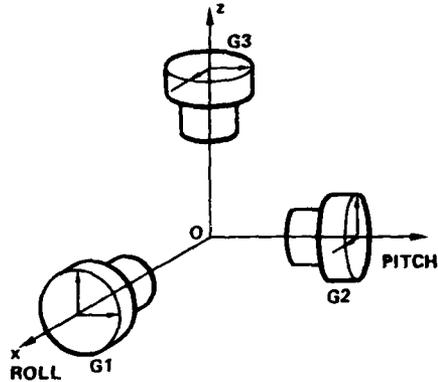
A, B = SCALE FACTORS

$$\begin{cases} U_x = A_x \Omega_x + B_x \Gamma_x \\ U_y = A_y \Omega_y + B_y \Gamma_y \end{cases}$$

SEVERAL POSSIBILITIES

EXAMPLE :

| | | |
|----|------------------|-----------------|
| G1 | $A_y = A_z = A$ | $B_y = B_z = 0$ |
| G2 | $A_z = A_x = A$ | $B_z = B_x = B$ |
| G3 | $A_x = A_y = -A$ | $B_x = B_y = B$ |

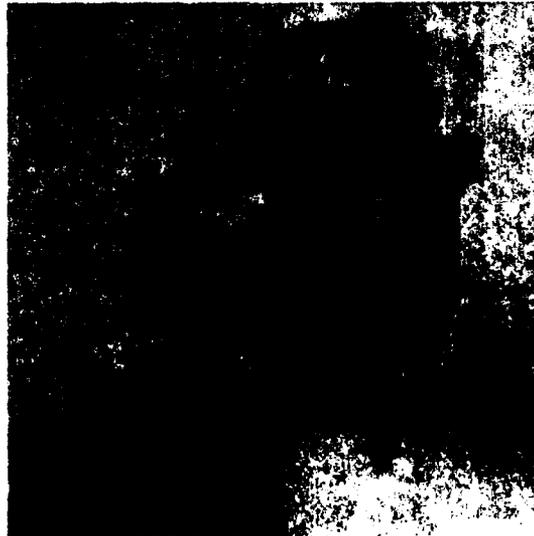


- **RELATION** ANGULAR RATE/LINEAR ACCELERATION

EXAMPLE : 1g → 10°/s



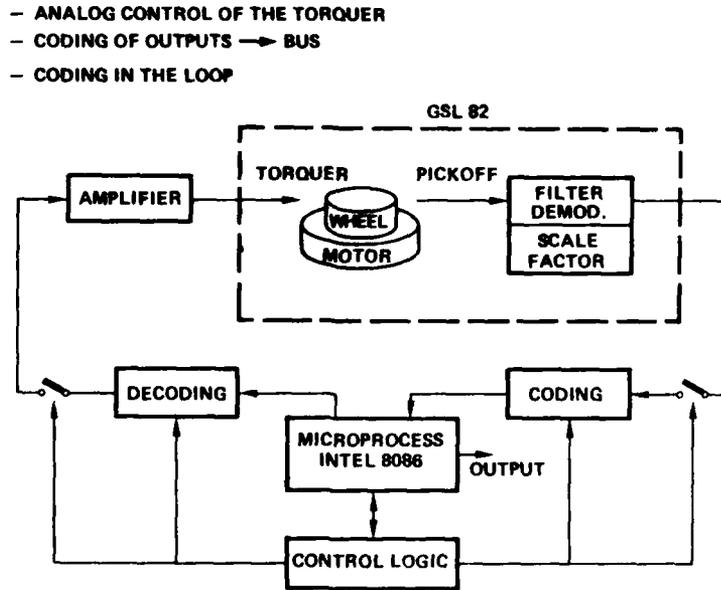
Figure 6 - COMPOSITE SENSOR GSL 82



| | | | | | |
|----------|---|------------------------------------|--------------------|---|--------------------|
| TYPE | = | DRY TUNED GYRO WITH MASS UNBALANCE | ANGULAR MOMENTUM | = | 5 000 DYNE CM SEC. |
| WEIGHT | = | 100 g | TORQUER YIELD | = | 400°/s per √WATT |
| DIAMETER | = | 30 mm | REACTION TIME | = | 1 SEC. |
| LENGTH | = | 35 mm | CONSUMPTION | = | 2 W |
| WHEEL | = | 6,5 g | THERMAL SENSOR | = | |
| SPEED | = | 320 Hz | SCALE FACTOR RATIO | = | 1g - 10°/s |



Figure 7 - CAGING LOOP



DECODING OFFSET → ERROR →
PERMANENT AUTOTEST AND COMPENSATION



Figure 8 - TRI-AXIS SYSTEM

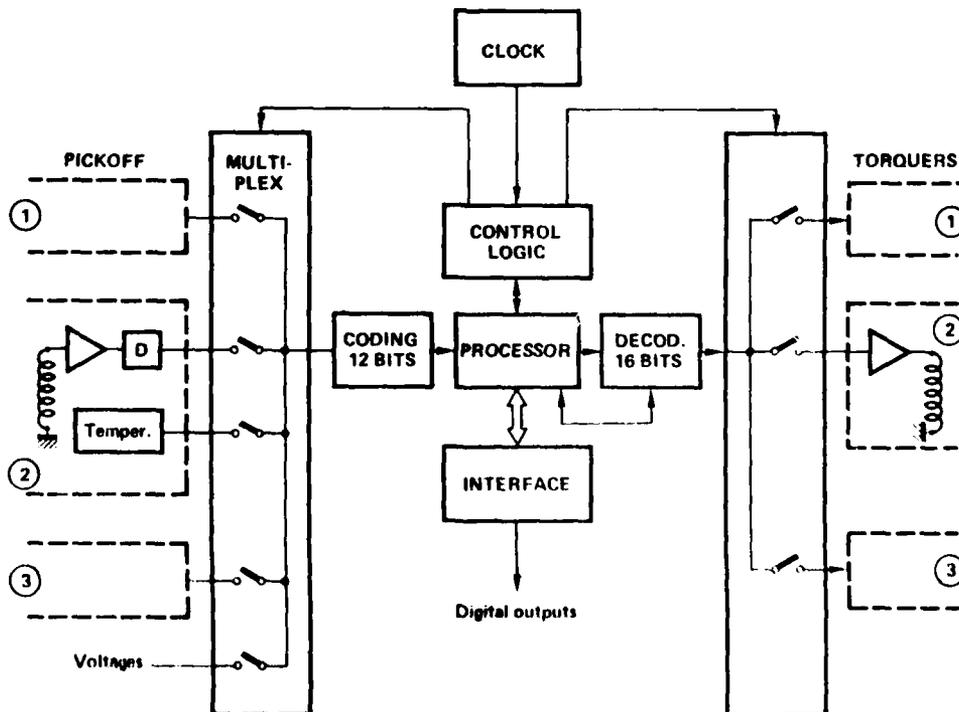




Figure 9 - PROCESSING UNIT

- ① CAGING LOOP
- ② COMPENSATIONS OF COMPOSITE MEASUREMENTS (VOLTAGE)
 - FIXED DRIFT
 - AXIS MISALIGNMENT
 - TRANSVERSE MASS UNBALANCE
 - ANGULAR MOMENTUM MODULATION
(OPTIONAL : GENERALLY NEGLIGIBLE)
 - SCALE FACTOR
 - SELF HEATING
 - STATIC TEMPERATURE ENVIRONMENT
 - DYNAMIC TEMPERATURE ENVIRONMENT
(OPTIONAL : GENERALLY NEGLIGIBLE)
 - BIAS OF THE LOOP
(CONTINUOUS MEASUREMENT)
- ③ COMPUTATION OF ROTATION RATE AND SPECIFIC FORCE COMPONENTS
- ④ CONTROL OF MULTIPLEX
- ⑤ CONTROL OF INTERFACE



Figure 10 - SIGNAL EXAMPLE
(STRAP DOWN GYRO - ACCELERO SYSTEM)

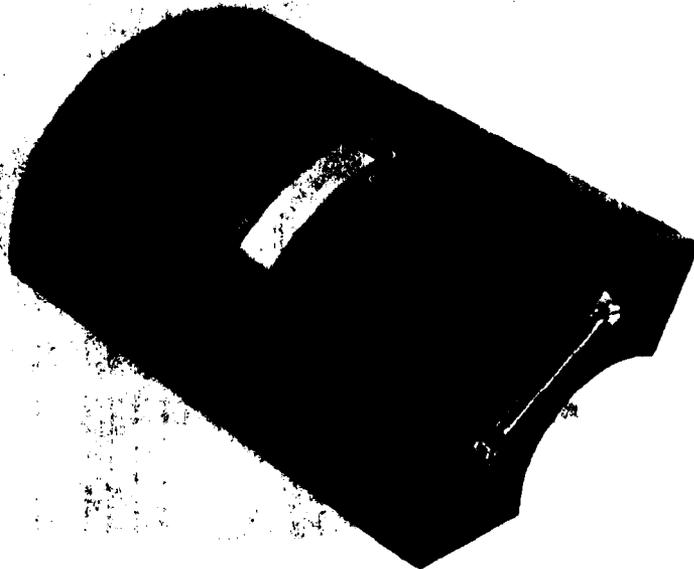




Figure 11 - ELECTRONICS
(CAGING AND PROCESSING)



Figure 12 - ELECTRONICS

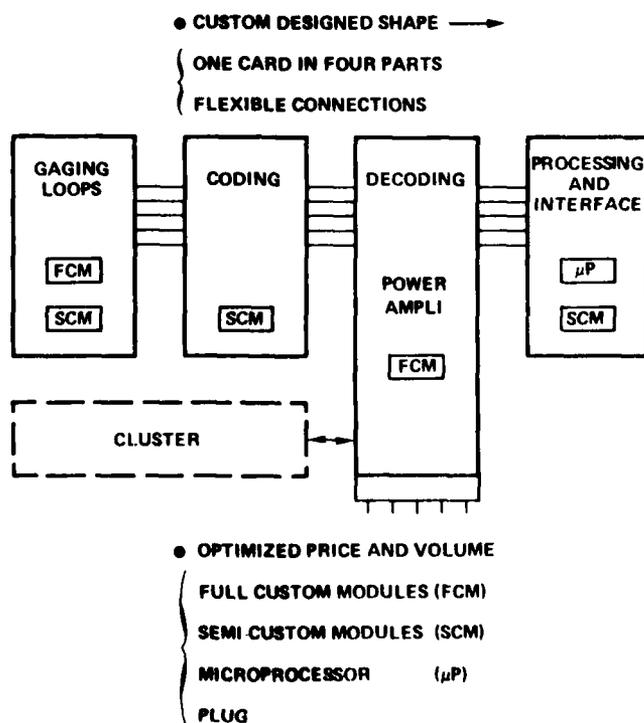




Figure 13 - SIGAL PERFORMANCES

ADAPTABILITY

● RANGES = { ANGULAR RATE
+
SIMULTANEOUS LINEAR ACCELERATION

→ RATIO { ANGULAR MOMENTUM
MASS UNBALANCE

● CRITICAL AXIS = ROLL ?

→ ORIENTATION OF MEASUREMENT AXIS



Figure 14 - SIGAL PERFORMANCES EX OF TYPICAL VALUES

- REACTION TIME < 2 SEC.

- RATE GYRO

RANGE : 500°/s

DRIFT { SHORT TERM : 5°/h
DAY/DAY : 40°/h

NON LINEARITY : 2.10⁻³

BAND WIDTH : 100 Hz

MISALIGNMENT : 5 ARC MIN.

- ACCELERO

RANGE : 50 g

BIAS { SHORT TERM : 1.5.10⁻⁴ g
DAY/DAY : 10⁻³ g

NON LINEARITY : 2.10⁻³

BAND WIDTH : 100 Hz

MISALIGNMENT : 5 ARC MIN.

- COUPLING

STABILITY : 2.10⁻⁴ g per °/s

- NOISE (0 - 250 Hz) : 0.5 °/s

- TORQUER

YIELD : 400 °/s per √WATT



Figure 15 - SIGAL DEVELOPMENT

SEVERAL PROTOTYPES

SPECIFIC { SHAPES
RANGES
ETC.

IN 1984

① LABORATORY TESTS
(STATIC AND DYNAMIC)

SOFTWARE { CALIBRATION
TEST
FLIGHT

② DYNAMIC QUALIFICATION
(FLIGHT SIMULATOR)

IN 1985

MISSILE FLIGHTS



Figure 16 - SIGAL CONCLUSIONS

- PROTOTYPE DEVELOPMENT OF SEVERAL MODELS FOR MID COURSE GUIDANCE AND CONTROL.
- CUSTOM DESIGNED SYSTEMS
 - . SHAPE
 - . VOLUME
 - . OPERATIONAL FLIGHT
 - . OPERATIONAL MODES

LASER GYROSCOPE RANDOM WALK DETERMINATION
USING A FAST FILTERING TECHNIQUE

by

John G. Mark and Alison Brown
Litton Guidance and Control Systems
5500 Canoga Avenue
Woodland Hills, CA 91367 USA

Abstract

Laser gyro performance is typically evaluated by measuring the random walk in angle associated with an instrument. However, the laser gyro outputs angle in quantized pulses and the angle quantization has a similar effect on gyro test data as white noise in angle.

In order to reduce the effect of quantization error on gyro test data, Litton has developed a high speed filter which outputs half second accumulated data samples. The quantization noise power in these samples is reduced by a factor greater than 512 while the random walk characteristics are virtually unaffected. High precision measurements of a gyro's random walk coefficient may therefore be obtained from a small data set.

This paper describes the operation of this fast filtering technique, analyzes its effect on gyro test data, and presents laser gyro test results which demonstrate the technique.

Laser Gyroscope Error Model

The stochastic errors in a laser gyro can be modeled by a random walk in angle, a white noise in angle and a quantization error. In Appendix A it is shown that quantization error has a Power Spectral Density identical to that of white noise in angle of spectral density $N_Q = \frac{Q^2}{12} T \text{ sec}^2/\text{Hz}$ where Q is the quantization level and T is the sample time interval. Hence for 0.5 sec

data and quantization $Q = 0.46 \widehat{\text{sec}}$, the spectral density of the equivalent white noise generated will be:

$$N_Q = (0.094 \widehat{\text{sec}}/\sqrt{\text{Hz}})^2.$$

The Root PSD of 0.5 sec simulated rate data from a gyro assumed to have a quantization error as above and a random walk spectral density of $N_R = (0.0015 \text{ deg}/\sqrt{\text{Hz}})^2$ is shown in figure 1. The dashed line represents the random walk level.

The Root PSD of gyro data with only random walk in angle is shown in figure 2. The variance of the rate estimate $\sigma_{\frac{\Delta\theta}{\Delta t}}^2$ is the area under the square of the Root PSD in figure 1.

$$\sigma_{\frac{\Delta\theta}{\Delta t}}^2 = \frac{2\sigma_Q^2}{T^2} + \frac{N_R}{T} \quad (1)$$

where σ_Q^2 is the quantization variance $\frac{Q^2}{12}$, $N_R = (0.09)^2 \widehat{\text{sec}}^2/\text{sec}$ is the random walk rate spectral density and $T = 0.5 \text{ sec}$ is the sample interval.

Figures 1 and 2 show how for 0.5 sec samples the quantization error dominates the measurement variance in Equation (1). If the sample time is increased or if the quantization level Q is decreased then the effect of the quantization error on the measurement variance is reduced. Increasing the sample time however, extends the test time as it takes longer to collect a representative number of samples.

Because of the presence of dither, bias, and secondary dither, the quantization error is randomized so that it has a similar frequency spectrum as white noise in angle. The quantization error can therefore be reduced by prefiltering the data at high frequencies. This will significantly reduce the

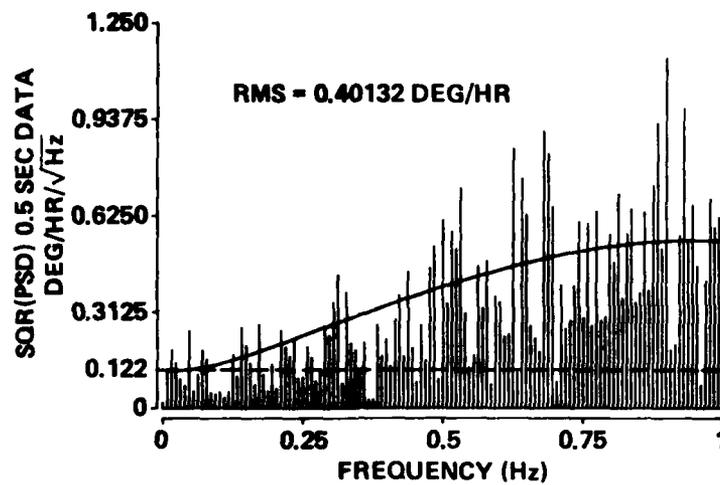


Figure 1. Root PSD of Simulated Gyro Rate Data With Unfiltered Random Walk and Quantization

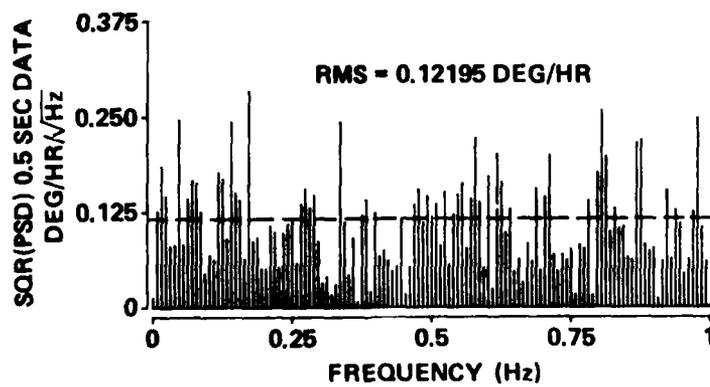


Figure 2. Root PSD of Simulated Gyro Data With Unfiltered Random Walk Only

power of the "white noise" generated by quantization error while only attenuating slightly the random walk rate noise variance σ_R^2 we are trying to observe.

Moving Average Pre-Filter

A 1,024 Hz moving average pre-filter is proposed. This filter consists of five cascaded moving average filters. The configuration used has several advantages. First, only additions and subtractions are required thereby minimizing necessary hardware while maximizing execution speed. Second, the filters have finite impulse response. That is, they achieve their final value within one sampling time. Third, the filters are "bit conservative" leading to exact arithmetic results with no truncation or rounding. Finally, although it is not immediately apparent, it is possible to implement this process with only minimal data storage.

It will be noted that only one filter is required to achieve the desired reduction of quantization noise. The additional filters are used to remove the sinusoidal component of dither.

The moving average filter is implemented as follows:

$$F_1(K) = F_1(K-1) + \Delta\theta(K) - \Delta\theta(K-2)$$

$$F_3(K) = F_3(K-1) + F_1(K) - F_1(K-8)$$

$$F_5(K) = F_5(K-1) + F_3(K) - F_3(K-32)$$

$$F_7(K) = F_7(K-1) + F_5(K) - F_5(K-128)$$

$$F_9(K) = F_9(K-1) + F_7(K) - F_7(K-512)$$

where $F_1(*)$, $F_3(*)$, $F_5(*)$, $F_7(*)$, and $F_9(*)$ are the five filter outputs and K is the index of the discrete data points.

The final output of the moving average pre-filter is then placed into an accumulator to provide 1/2 second samples.

$$\Delta\theta_F = 2^{-25} \sum_{K=1}^{512} F_9(K)$$

The overall transfer function of the filter is given by:

$$F(Z) = 2^{-25} \frac{1-Z^{-512}}{1-Z^{-1}} \prod_{M=0}^4 \frac{1-Z^{-2^{(2M+1)}}}{1-Z^{-1}} \quad (2)$$

Figures 3a and 3b show the filter transfer function and response.

Attenuation is over 25 dB at 2 Hz and increases progressively at higher frequencies. Over 200 dB of attenuation is achieved at dither frequencies (~400 Hz) so that dither is essentially removed from the output $\Delta\theta_F$.

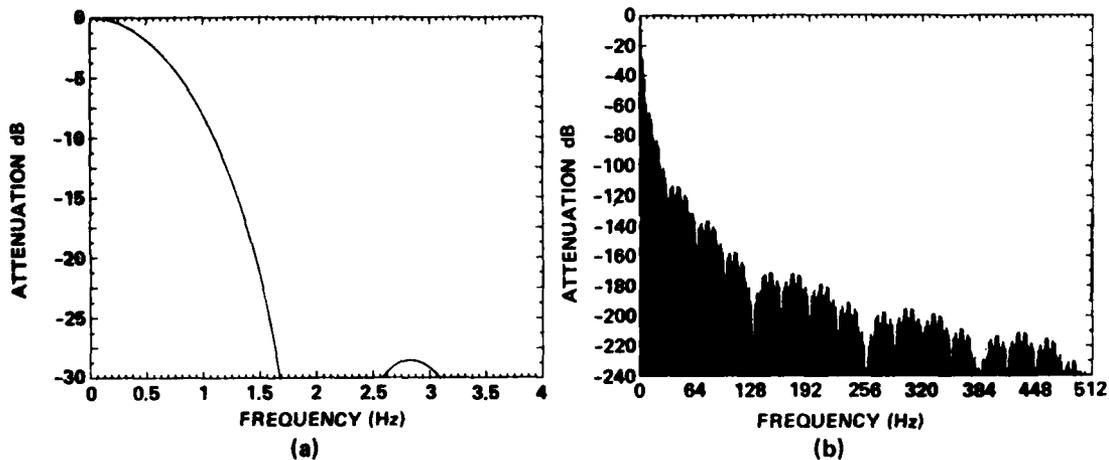


Figure 3. Moving Average Prefilter Transfer Function

In Appendix B, the effect of the moving average filter and accumulator on random walk in angle and "white noise" in angle produced by quantization is derived. The reduction of the random walk rate variance is predominantly due to the longest summation in the moving average filter which is of the form:

$$y(K) = y(K-1) + X(K) - X(K-512).$$

If the accumulator sums over $m = 512$ samples, then, from Appendix B, the RMS of the filtered 1/2 second rate data $\frac{\Delta\theta_F}{\Delta t}$ is approximately given by

$$\frac{\sigma_{\Delta\theta_F}^2}{\Delta t} \approx \frac{2\sigma_Q^2}{512T^2} + \frac{N_R}{T} \left(1 - \frac{1}{3}\right) \quad (3)$$

Simulation Results

To illustrate the performance of the moving average pre-filter, gyro data was simulated for a laser gyroscope with 0.46 sec quantization and a random walk in angle of 0.0015 deg/ $\sqrt{\text{Hr}}$. Simulation data was generated as shown in figure 4 to demonstrate the effect of the filter on quantization error alone and on a combination of quantization error and random walk.

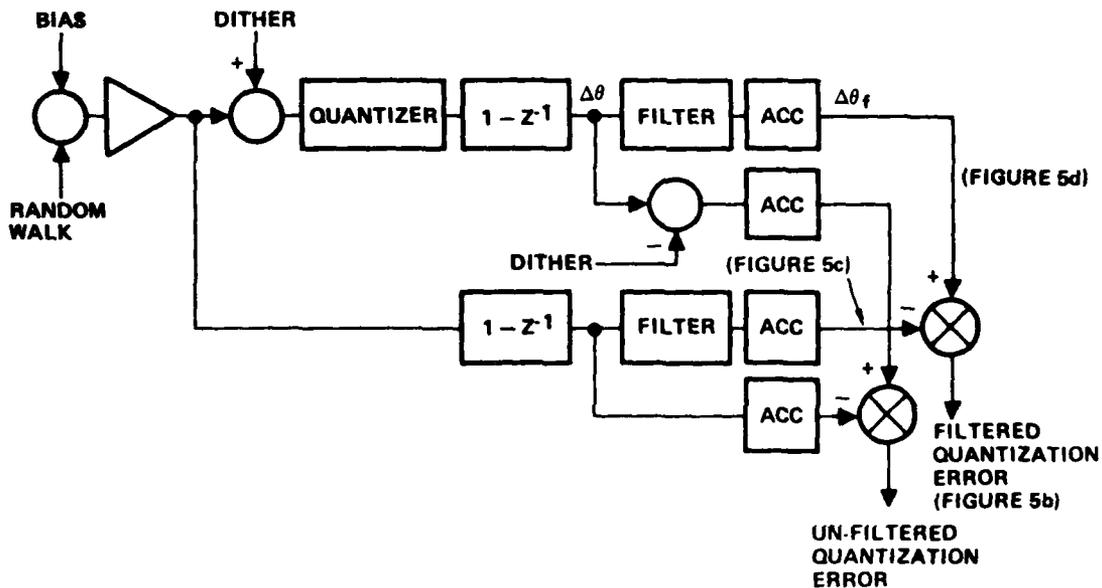


Figure 4. Model of Laser Gyroscope Simulation

Figures 5a, b, c, and d show the simulated data. The same data set was used throughout the simulations to generate PSD's. Figure 5a is a 256 data point sample of unfiltered random walk. Figure 5b shows the filtered quantization. The data in figure 5c represents the filtered random walk data. Comparison with figure 5a shows that while the RMS of the random walk is slightly reduced by the filter, the data is virtually unchanged. Finally, figure 5d shows the filtered sum of quantization and random walk. It is clear that filtering has reduced quantization noise to a level well below random walk.

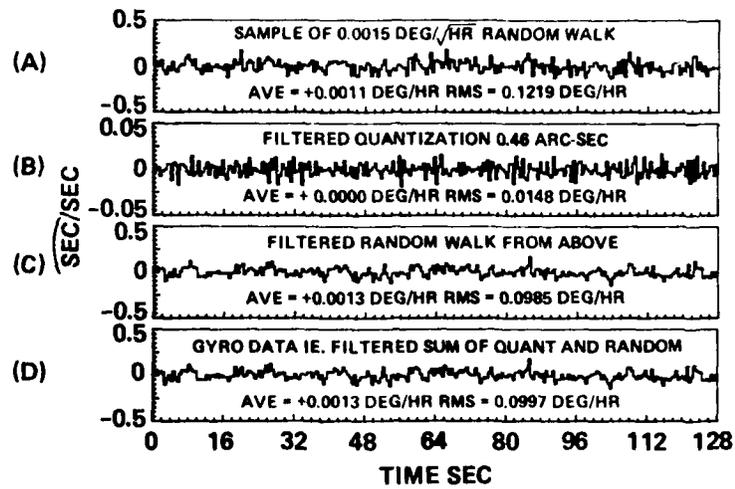


Figure 5. Filtered Simulation Data

Figure 6 shows a plot of the single-sided Root PSD of the quantization error on the unfiltered simulated gyro data $\frac{\Delta\theta}{\Delta t}$.

From Appendix A this Root PSD should fit the curve

$$\sqrt{2} \left(\frac{4\sigma_Q^2}{T} \sin^2 \pi f T \right)^{1/2}$$

which is also plotted on figure 6. The expected RMS of the data is

$$\hat{\sigma}^2 = \frac{2\sigma_Q^2}{T^2}$$

or $\hat{\sigma} = 0.376$ deg/hr which is close to the actual RMS of the simulation data, $\sigma = 0.395$ deg/hr. In figure 7 the filtered quantization error is shown. Theory predicted that the RMS error should have been reduced by the Moving Average Filter by $\sqrt{512}$ to 0.0166 deg/hr. The filter actually reduced the RMS to 0.0148 deg/hr, the additional attenuation being due to the additive effects of the additional filters which were not accounted for by equation (3).

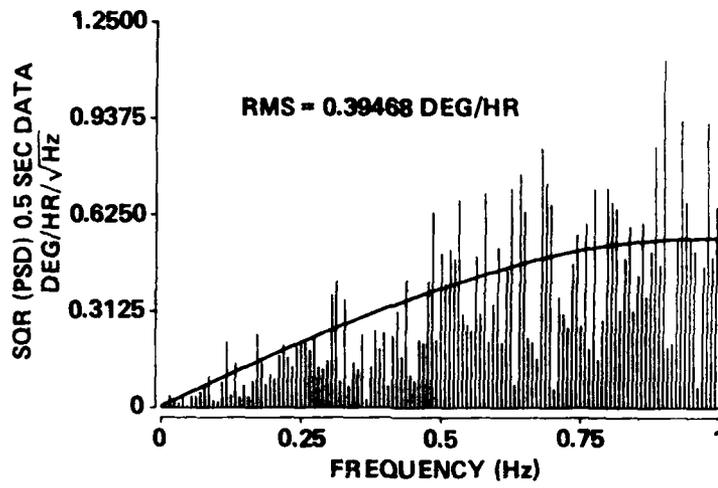


Figure 6. Root PSD of Unfiltered Quantization Error

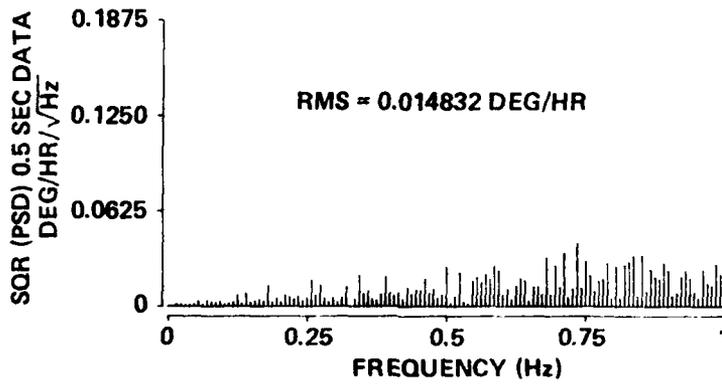


Figure 7. Root PSD of Filtered Quantization Error

Figure 1 shows the Root PSD of the unfiltered gyro data containing both gyro random walk, which is to be estimated, and quantization error. The Root PSD should fit the curve

$$\sqrt{2} \left(N_R + \frac{4\sigma_Q^2}{T} \sin^2 \Pi ft \right)^{1/2}$$

which is also plotted on figure 1. The RMS of the data can be calculated from

$$\frac{\sigma_{\Delta\theta}^2}{\Delta t} = \frac{2\sigma_Q^2}{T^2} + \frac{N_R}{T}$$

or $\frac{\sigma_{\Delta\theta}}{\Delta t} = 0.40$ deg/hr which was as calculated from the data. The filtered gyro data is shown in figure 8. As can be seen this Root PSD closely approximates that of a random walk.

The RMS calculated from this data is $\frac{\sigma_{\Delta\theta_F}}{\Delta t} = 0.0997$ deg/hr. Working backwards through equation (3) and converting units gives:

$$\hat{N}_R = \left(\left(1 - \frac{1}{3}\right)^{-1/2} \sqrt{T} \frac{\sigma_{\Delta\theta_F}}{\Delta t} \right)^2$$

$$\hat{N}_R = (0.0863 \text{ Deg/Hr} / \sqrt{\text{Hz}})^2$$

$$\hat{N}_R = (0.00144 \text{ deg}/\sqrt{\text{Hr}})^2$$

This is to be compared to the value of $N_R = (0.0015 \text{ deg}/\sqrt{\text{Hr}})^2$ which was used to generate the simulation. Therefore, using the filtered gyro data to calculate the random walk has introduced no significant errors to the estimate and has allowed a much faster determination of random walk than would have been possible with conventional methods.

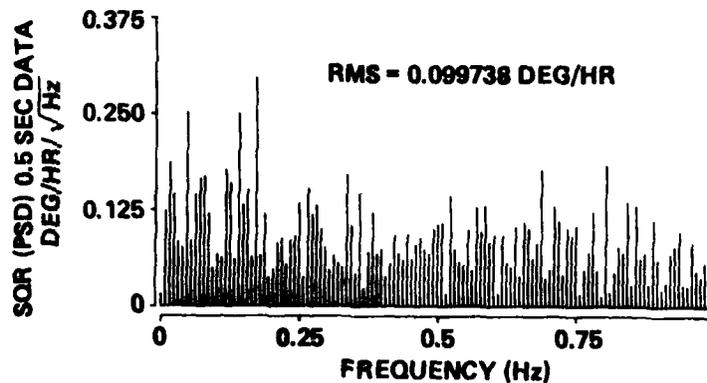


Figure 8. Root PSD of Simulated Filtered Gyroscope Data

SUMMARY OF GYRO DATA

Table I summarizes data collected on several gyros in a comparison test. The results indicate good correlation between the random walk estimated from 100 sec samples using auto-variance in a standard ATP and those using samples which have been processed by the Moving Average filter.

Conclusion

This moving average filter has been used at Litton for the past year to evaluate many Ring Laser Gyros. Filtered data samples have given reliable random walk coefficients from the RMS of as few as 256 samples (128 secs) when compared with unfiltered data taken over much longer times (as high as 24 hours).

TABLE I.

| Gyro Serial No. | Fast Filter Random Walk Estimation Deg/ $\sqrt{\text{Hr}}$ | Random Walk From ATP Data Deg/ $\sqrt{\text{Hr}}$ |
|--------------------|--|---|
| 275 | 0.0026 | 0.0031 |
| 1232 | 0.0020 | 0.0020 |
| 959 | 0.0011 | 0.0009 |
| 1088 | 0.0010 | 0.0011 |
| 1049 | 0.0009 | 0.0009 |
| 470 | 0.00130 | 0.0018 |

Simulated data has shown that with 256 half second samples, the Moving Average Filter estimated a random walk coefficient of 0.00144 deg/ $\sqrt{\text{Hr}}$ which was very close to the random walk simulated.

REFERENCES

1. "Fundamental Limit of Ring Laser Gyro", T. Callaghan, S. Callaghan, J. Hause, C. Tetteimer, F. Aranowitz, presented at Gyro Symposium, Stuttgart, September 1982.
2. "The Fast Fourier Transform", E. Oran Brigham, Prentice Hall.
3. "Quantization Reduction for Evaluating Laser Gyro Performance Using a Moving Average Filter", J. Mark, A. Brown, and T. Matthews, presented at IEEE ICASSP, San Diego, CA, March 1984.

APPENDIX A

Effect of Quantization Error on Angle

The quantization error $\tilde{\theta}_k$ can be modeled as

$$\tilde{\theta}_k = \hat{\theta}_k - \theta_k = Q_0 + Q_k$$

where Q_0 is the initial quantization error and Q_k is the error introduced on the k th angle count.

Both errors are uniformly distributed $\pm 1/2$ pulse and so have variance $\sigma_Q^2 = Q^2/12$ where Q is the quantization level.

The auto-correlation function of $\tilde{\theta}$, the angle error is given by:

$$\gamma_{\theta}(k) = E \{ \tilde{\theta}_j \tilde{\theta}_{j+k} \} = \begin{cases} 2\sigma_Q^2; k=0 \\ \sigma_Q^2; k \neq 0 \end{cases}$$

This is equivalent to the auto-correlation function of a bias with variance σ_Q^2 and white noise with spectral density $N_Q = T\sigma_Q^2 \text{ sec}^2/\text{Hz}$ where T is the sample time.

Incremental angle measurements are generated using the equation

$$\Delta\theta_k = \theta_k - \theta_{k-1}$$

so the auto-correlation function of $\Delta\theta_k$ is

$$\begin{aligned} \gamma_{\Delta\theta}(0) &= E \left[\Delta\tilde{\theta}_k \Delta\tilde{\theta}_k \right] = E \left[\tilde{\theta}_k^2 \right] - 2E \left[\tilde{\theta}_k \tilde{\theta}_{k-1} \right] \\ &\quad + E \left[\tilde{\theta}_{k-1}^2 \right] = 2\sigma_Q^2 \\ \gamma_{\Delta\theta}(1) &= \gamma_{\Delta\theta}(-1) = E \left[\Delta\tilde{\theta}_k \Delta\tilde{\theta}_{k-1} \right] \\ &= E \left[(\tilde{\theta}_k - \tilde{\theta}_{k-1}) (\tilde{\theta}_{k-1} - \tilde{\theta}_{k-2}) \right] = -\sigma_Q^2 \\ \gamma_{\Delta\theta}(K) &= 0; K \neq -1, 0, 1 \end{aligned}$$

The Fourier Transform of $\gamma_{\Delta\theta}(K) \phi \left(\frac{n}{N\Delta t} \right)$ gives the PSD for the incremental rate measurements $\frac{\Delta\theta}{\Delta t}$ with quantization error.

$$\begin{aligned} \phi \frac{\Delta\theta}{\Delta t} \left(\frac{n}{N\Delta t} \right) &= \frac{1}{T} \sum_{k=0}^{N-1} \gamma_{\Delta\theta}(k) e^{-j2\pi nk/N} \\ &= 4 \frac{\sigma_Q^2}{T} \left(\sin \frac{\pi n}{N} \right)^2 \end{aligned}$$

For a white noise process, the PSD could also be obtained from the amplitude response as:

$$\begin{aligned} \phi \frac{\Delta\theta}{\Delta t} \left(\frac{n}{N\Delta t} \right) &= \frac{\sigma_Q^2}{T} \times (\text{Amplitude Response}) = \frac{\sigma_Q^2}{T} \left| 1 - z^{-1} \right|^2 \\ &= \frac{\sigma_Q^2}{T} \left(2 - 2 \cos \left(\frac{2\pi n}{N} \right) \right) = \frac{4\sigma_Q^2}{T} \sin^2 \left(\frac{\pi n}{N} \right) \end{aligned}$$

APPENDIX B

Effect of Moving Average Filter on Gyro Data

The noise acting on a gyroscope, white noise in angle (due mainly to quantization) and white noise in rate (random walk in angle) can be modeled as shown in figure B-1.

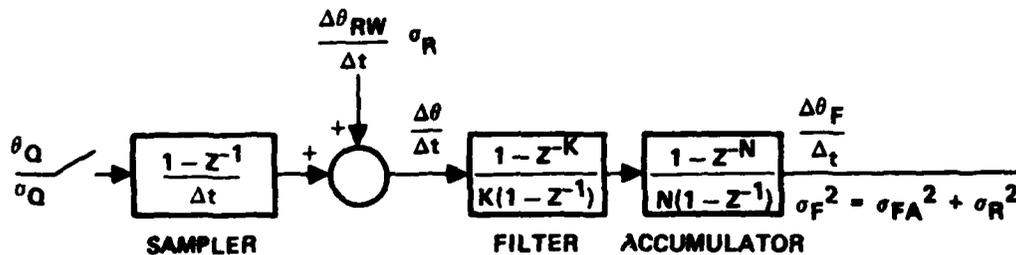


Figure B-1. Model of Laser Gyro and Prefilter Driven by Random Walk and Quantization Error

If the white noise due to quantization has variance σ_Q^2 , then the rate variance of the filtered output σ_{fA}^2 (considering only one element of the moving average filter) is given by:

$$\sigma_{fA}^2 = \frac{\sigma_Q^2}{2\pi} \int_{-\pi}^{\pi} F_A(\phi) F_A^*(\phi) d\phi$$

where

$$F_A(\phi) = \frac{1-Z^{-K}}{K\Delta t} \frac{1-Z^{-N}}{N(1-Z^{-1})} \quad \text{is the filtering function.}$$

$$\sigma_{fA}^2 = \frac{\sigma_Q^2}{2\pi N^2 K^2 (\Delta t)^2} \int_{-\pi}^{\pi} \{K + 2(K-1) \cos \phi + \dots + 2 \cos(K-1)\phi\} (2-2 \cos N\phi) d\phi$$

if $K < N$ then only $2K$ survives in the integral.

Thus,

$$\sigma_{fA}^2 = \frac{2\sigma_Q^2}{N^2(\Delta t)^2 K}$$

Because the filter has a finite output response (i.e., an input signal will be output within one sampling time), the filtered quantization will have a PSD function of the same form as the unfiltered quantization.

The accumulator alone will reduce the variance to $\frac{2\sigma_Q^2}{N^2(\Delta t)^2}$ so the addition of the Kth section of the moving average filter has reduced the variance of the output due to white noise in angle by a further $1/K$. Adding in the other sections will further reduce the variance but the most significant effect is introduced by the section with the greatest delay, K.

If the random walk in angle has spectral density N_R , then the additive effect of the variance of $\frac{\Delta\theta}{\Delta t}$ is $\sigma_R^2 = \frac{N_R}{T}$. The filtered rate variance due to random walk in angle σ_{fR} is given by

$$\begin{aligned} \sigma_{fR}^2 &= \frac{\sigma_R^2}{2\pi K^2 N^2} \int_{-\pi}^{\pi} \frac{(1-Z^{-K})}{(1-Z^{-1})} \frac{(1-Z^{-N})}{(1-Z^{-1})} \{ \dots \}^* d\phi \\ &= \frac{\sigma_R^2}{2\pi N^2 K^2} \int_{-\pi}^{\pi} \left\{ K + \sum_{j=1}^{K-1} 2(K-j) \cos j\phi \right\} \left\{ N + \sum_{j=1}^{N-1} 2(N-j) \cos j\phi \right\} d\phi \end{aligned}$$

Note that only cosine² terms survive, i.e.,

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \cos^2 j\phi d\phi = 1/2$$

Thus,

$$\sigma_{fR}^2 = \frac{\sigma_R^2}{K^2 N^2} \left(KN + 2 \sum_{j=1}^{K-1} (K-j)(N-j) \right)$$

which can be summed exactly giving

$$\sigma_{fR}^2 = \frac{\sigma_R^2}{N} \left\{ 1 - \frac{K^2 - 1}{3NK} \right\}$$

Again, as other sections are added, the variance is reduced further but the most significant effect is introduced by the section of the filter with largest K. If K = 512 and N = 512 then T = NΔt = 0.5 sec and the total filtered rate variance is:

$$\sigma_{\frac{\Delta\theta_f}{\Delta t}}^2 = \frac{2\sigma_Q^2}{512(N\Delta t)^2} + \frac{N_R}{N\Delta t} \left\{ 1 - \frac{512^2 - 1}{3(512)(512)} \right\}$$

$$\sigma_{\frac{\Delta\theta_f}{\Delta t}}^2 \approx \frac{2\sigma_Q^2}{512T^2} + \frac{N_R}{T} \left\{ 1 - \frac{1}{3} \right\}$$

It is seen that the quantization term is reduced considerably more than the random walk term to the point where

$$\sigma_{\frac{\Delta\theta_f}{\Delta t}}^2 \approx \frac{2}{3} \frac{N_R}{T}$$

The PSD of $\frac{\Delta\theta_f}{\Delta t}$ may now be easily formulated. As explained earlier, filtered quantization has a PSD of the same form as unfiltered quantization. Again invoking the finite output response of the filter, we can determine that the random walk autocorrelation function can have only a DC component and one

additional symmetric component. Further, the filter gain at DC must be unity since a constant input into the filter will yield the same constant output.

From the previous calculation, we find that the DC term is approximately $1 - 1/3 = 2/3$. Therefore, the PSD of the rate is given by:

$$\begin{aligned} \frac{\phi_{\Delta\theta_F}}{\Delta t} \left(\frac{n}{N\Delta t} \right) &= \frac{4\sigma_Q^2}{(N\Delta t)} \sin^2 \left(\frac{\pi n}{N} \right) + N_R \left\{ \frac{2}{3} + \frac{1}{3} \cos \left(\frac{2\pi n}{N} \right) \right\} \\ &= \frac{4\sigma_Q^2}{T} \sin^2 \left(\frac{\pi n}{N} \right) + N_R \left[1 - \frac{2}{3} \sin^2 \left(\frac{\pi n}{N} \right) \right] \end{aligned}$$

THE SERIES 2000 - A MINIATURE GAS BEARING DTG
FOR LOW COST STRAPDOWN GUIDANCE AND CONTROL

by
Dr. G. Beardmore
SMITHS INDUSTRIES AEROSPACE & DEFENCE SYSTEMS
Integrated Systems Group
Bishops Cleeve,
Cheltenham, Glos. GL52 4SF
United Kingdom

SUMMARY

The development and status of a novel Dynamically Tuned Gyroscope (DTG) is described, with particular reference to the specialised technology employed to yield a cost effective design. Its unique configuration possesses a number of important advantages over those used in conventional tuned sensors, and allows, for the first time, a self-acting gas bearing to be incorporated in a strapdown DTG. Other features of the design include the use of a capacitive pick-off to replace the traditional inductive version, a high performance spin motor and a fully fabricated flexure hinge. This hinge overcomes many of the technical and commercial disadvantages of the now familiar 'carved-from-solid' hinges. The paper discusses the problems that have hitherto prevented the application of gas bearing technology to tuned sensors and describes how these have been overcome by the Series 2000 design. Current performance of the sensor is summarised and its advantages over contemporary rate sensors are compared with the requirements of future guidance and control applications. These include airborne, missile, underwater and surface systems where life, reliability, ready time and affordability are of prime importance.

1. INTRODUCTION

Smiths Industries have manufactured a range of single axis rate sensors at their Cheltenham facility for over forty-five years. The majority of these sensors have utilised ball bearing technology but over the last 20 years, a range of miniature open loop gyros have been developed in which the ball bearings have been entirely replaced by gas bearings and flexure suspensions. The introduction of this technology has resulted in an impressive improvement in life and reliability, with in-service MTBFs of over 85,000 hours currently being achieved [1].

In 1979, a requirement was identified for a low cost two axis rate sensor for application in a wide range of current and future strapdown systems up to and including IN quality. The Dynamically Tuned Gyroscope was a natural candidate for this role because its principles were well proven and the concept of dynamic tuning was accepted throughout the industry. However, a preliminary study revealed some serious limitations with existing DTG technology in terms of both manufacturing cost and operational life, and showed that these limitations were directly attributable to certain common features which included the flexure hinge, the ball bearings and the general configuration.

The design philosophy therefore centered around the need to eliminate the spin axis ball bearings and to introduce a self-acting gas bearing to support the rotating flexure hinge and rotor assembly. This in turn dictated a fairly radical re-appraisal of the traditional rotor/flexure hinge design which invariably overhangs the central flange and bearing assembly, and in which the hinge is normally carved from solid metal. In addition to these specific technical considerations, there was an obvious advantage in utilising the existing and proven gas bearing and flexure pivot technology already employed throughout the company's Series 700 range of miniature gyroscopes [2].

2. DEVELOPMENT HISTORY

Research and Development of the Series 2000 DTG, Fig.1, commenced in 1980 and took the form of a joint venture programme between Smiths Industries and the UK Ministry of Defence. The programme was carried out in two phases, culminating in the manufacture of both breadboard and prototype hardware for joint evaluation by both parties. A dedicated team was established at Cheltenham to carry out the design, development and manufacturing aspects of the programme and close liaison was maintained throughout with the Radio and Navigation Department of the Royal Aircraft Establishment at Farnborough. Independent evaluation of the gyro and its associated electronics has been carried out by the Radio and Navigation Department at Farnborough and evaluation of the latest improved prototype hardware is continuing.

3. GAS BEARING vs DTG

Design of a gas bearing for application in a DTG involves a basic conflict in requirements and it is perhaps significant that no previous attempt appears to have been made to combine these otherwise proven technologies. The gas bearing utilises the viscous effects of the lubricating gas to generate load carrying capacity and, by definition, requires sufficient gas to be present in order to establish reliable hydrodynamic lubrication. Conversely, the DTG performance is degraded by the action of gas damping torques on its inertial elements and all contemporary sensors of this type

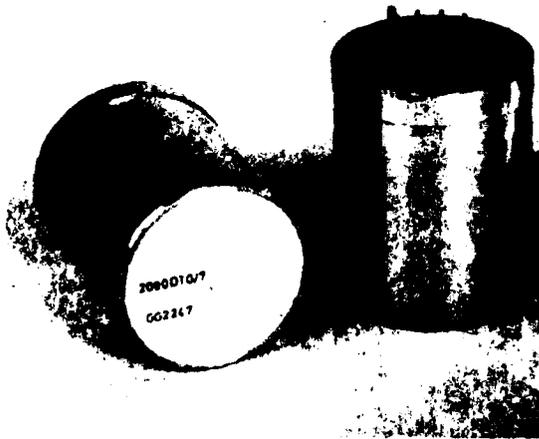


FIG 1 THE SERIES 2000 DTG

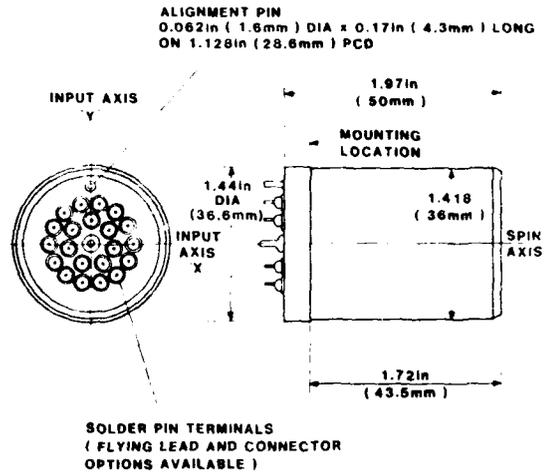


FIG 2 OUTLINE OF THE SERIES 2000 DTG

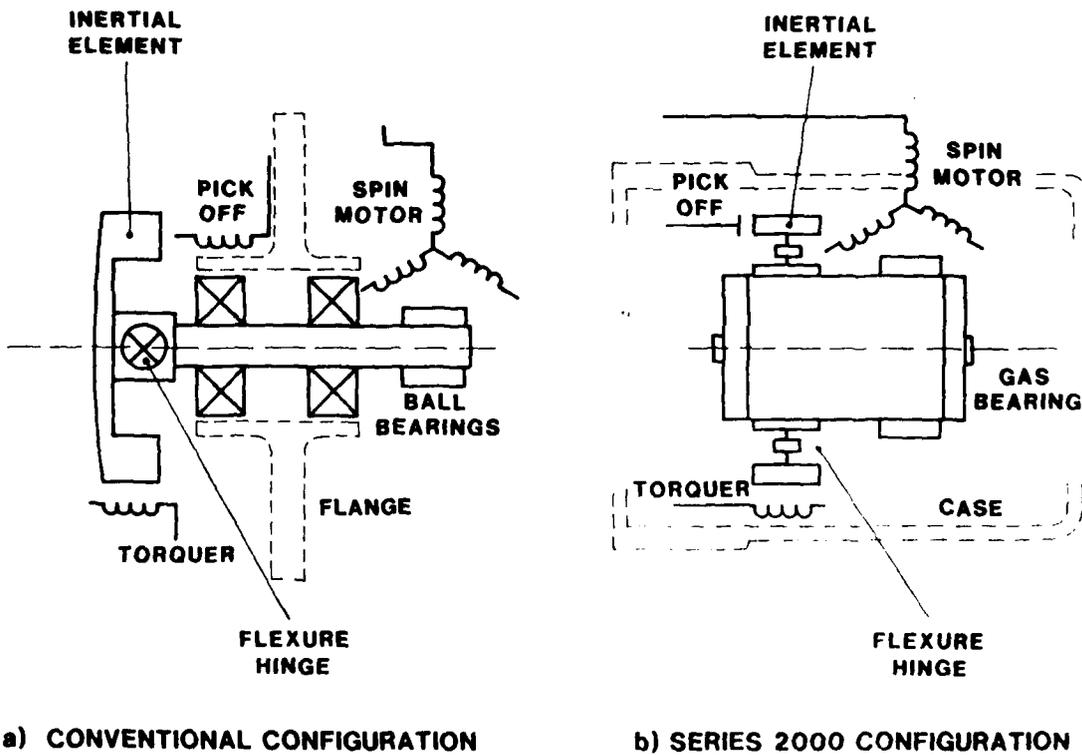


FIG 3 COMPARISON OF CONVENTIONAL AND SERIES 2000 DTG CONFIGURATION

are obliged to operate at a reduced case pressure in order to achieve acceptable drift rates. The magnitude of the effect depends upon the geometry and gas involved but case pressure reductions down to one quarter atmosphere are typical. This lower limit is itself a compromise in order to prevent deterioration or loss of the lubricant in the ball bearings. At first sight, these conflicting requirements appear to preclude the application of gas bearings to dynamically tuned sensors. The problem is compounded when one recognises the implications of the fact that the inertial assembly overhangs the bearing system in a conventional DTG design. This overhang represents a difficult design problem for any bearing system but for a self-acting gas bearing it presents almost insurmountable difficulties.

To overcome these conflicts, the problem was approached from two directions. First, using experience gained on the Series 700 sensors, a gas bearing was designed to operate at reduced ambient pressures. Secondly, a DTG design that would operate successfully at relatively high case pressures was evolved by careful aerodynamic design of the inertial element and its immediate surroundings. Most important of all, bearing overhang was eliminated by the design of a novel fabricated flexure hinge which allows the gas bearing to pass completely through its centre of mass. By combining these features it has been possible to produce a gas bearing DTG that will operate over a wide range of case pressures from sub to super atmospheric. In fact, an environmental/performance trade-off can be achieved by simply selecting the case pressure on any given mechanical design and the manufacturing advantages of this arrangement are obvious. Furthermore, the Series 2000 bearing is specifically designed to use ordinary clean room air as a lubricant, in common with our existing range of gas bearings and in contrast to those used in other gas bearing sensors. The use of air as opposed to an inert or rare gas fill has considerable benefits in terms of ease of manufacture and, contrary to what might be expected, results in a more stable surface chemistry at the gas bearing surfaces. It also has unique benefits in relation to the Boron Carbide selected for the bearing components, as will be discussed.

4. CONFIGURATION

The conventional layout of a DTG is illustrated schematically in Figure 3.a. Design details vary somewhat but the basic configuration has changed little over the past twenty years. The overhang of the flexure hinge does not contribute to long bearing life and any sag of the central shaft in its bearings is automatically sensed by the pick-off and appears as an error output. This arrangement lacks symmetry and the thermal paths of the torquers and spin motor are long and are not necessarily independent.

In contrast, the configuration of the Series 2000 DTG is shown in Figure 3.b. The overhang on the spin bearing has been entirely eliminated and the design has gained in both symmetry and simplicity. A relatively large gas bearing has been inserted through the mass centre of the flexure hinge assembly and both the hinge and the spin motor driving ring lie between the thrust bearings on either end of the journal bearing. Both torquers and the spin motor are heat sunk directly to the outer case, which also provides magnetic screening and a hermetic seal. This multi-function case also serves to locate the three modular sub-assemblies and therefore replaces both the central flange and the screening covers on contemporary designs. The one-time build around a central flange is replaced by three fully interchangeable modules, each of which can be tested independently prior to final assembly and the combined operation of the three modules can be verified functionally prior to inserting them into the case.

The case and mounting arrangements are designed to allow the user to take the fullest possible advantage of the internal heat sinking and the DTG can therefore be embedded into a mounting block and secured against the narrow mounting location with three synchro style clamps. This preferred method of mounting allows external heat sinking along the entire length of the case and reduces thermal gradients to a minimum. Alternatively the sensor can be mounted on a flat plate or bulkhead in the normal manner. Where multi-axis packaging is required, the aspect ratio of the DTG is important as sensors must be mounted with their spin axes at right angles and it is the overall size of the cluster that is important rather than the case diameter alone. The Series 2000 has been designed with this in mind and makes maximum use of the volume available while minimising the overall size of the cluster. The input axes are identified by a case label and, more accurately, by an alignment pin at the terminal end of the case.

5. GAS BEARING DESIGN

An H-Form or spool bearing was eventually selected as the most appropriate configuration for this demanding application. Other arrangements have their advocates but an H-Form bearing offers the greatest design flexibility and is easier to manufacture than conical or hemispherical systems [3]. Logarithmic spiral pumping grooves are cut into each stationary thrust surface and helical grooving is used on the journal to suppress half speed whirl and to enhance the pumping action.

Designing a bearing which would operate satisfactorily at both low and high pressures proved to be a difficult task. In conventional gyroscopes, the wheel can be operated at speeds of 24,000 RPM and above, but for tunnel sensors, other constraints generally limit the rotational speed to about half this value. This reduction in speed, together with the reduced ambient pressure and the need to maintain a precisely defined spin axis for the flexure hinge effectively reduces the number of variables readily

available to the designer and more subtle techniques must be employed. Operation at low pressures and clearances brings with it an increased risk of performance degradation due to slip flow effects (where there are simply not enough gas molecules for the gas to behave as a gas) while bearings optimised for operation under these conditions may consume excessive power at higher pressures. Finite difference analysis had been used to optimise our previous gyro bearings and, based upon this experience, a technique was evolved which allowed the design of a DTG gas bearing to proceed. Similarly, the geometric design of the assembly was studied with a view to achieving a low and reasonably balanced specific loading on both journal and thrust elements. The final running clearances selected were below 40 micro inches (1 micron) and some measure of the combined efficiency of the bearing and its drive can be gained by noting that the total synchronous power consumption is less than one watt at a case pressure of one atmosphere.

All the bearing parts, Fig.5, are machined from solid Boron Carbide. This is a hot pressed ceramic material with outstanding physical and chemical properties. It is the next hardest bulk solid to diamond (3000 VPN), is lighter than aluminium (2.52 g/cc) and has excellent wear and friction characteristics. Unlike most other gas bearing materials it does not require a boundary lubricant to achieve an acceptable start/stop life because under rubbing conditions, a combination of local frictional heating and oxygen produce a soft dry lubricant (boric oxide) in microscopic quantities at the point of contact. Boron Carbide has been used extensively in Smiths Industries gas bearings for many years and bearings incorporating this material have demonstrated over 100,000 start/stop cycles without failure. By using the same material for all the bearing components, differential expansion problems are avoided but exceptionally tight machining tolerances are still required in order to preserve the running clearances. All parts are machined to micro-inch tolerances and the design allows the gas bearing to be assembled and tested prior to final assembly of the gyro.

6. SPIN MOTOR

A six pole, three phase hysteresis motor is used to drive the bearing and flexure assembly at its nominal tuned speed (N) of 200 Hz. The choice of poles and phases was influenced by the need to minimise 2N frequencies which could interact with the flexure assembly and cause drift errors. Motor design was based upon the successful Series 700 spin motor and has been optimised to meet the special requirements of a hydrodynamic bearing. These include the need to generate a high instantaneous starting torque, while maintaining efficient operation under synchronous conditions. The Series 2000 motor differs from traditional hysteresis motors insofar as it has an integral fringing band (flux bridge) across the teeth of the laminations and uses a solid, as opposed to a laminated, hysteresis ring. Under starting conditions, most of the torque is generated by inductive currents flowing in the solid driving ring and the motor only assumes the characteristics of a true hysteresis machine at synchronism. No ligaments are necessary and the power leads are fed through the centre of the gas bearing to keep them well clear of the sensitive inertial element. The spin motor module incorporates a magnetic shield and is thermally bonded to the outer case.

7. FABRICATED FLEXURE ASSEMBLY

The now traditional method of forming the flexure hinge is to start with a solid block of metal and carve it away by grinding and spark erosion techniques to establish a two axis Hookes Joint. This method of manufacture has many disadvantages. Grain flow cannot be aligned along the length of each flexure and local defects coincident with the neck of the flexure can result in an unacceptable scrap rate. Electro-discharge machining (EDM) leaves a poor granular finish and even if this is locally polished, sub-surface damage remains and the flexure is weakened. The design must allow access to all areas for machining and the final tolerance build-up is normally high. Since spring rate is proportional to the third power of the flexure thickness, the rate, and hence the tuned speed can vary in practice by up to $\pm 30\%$. Mechanical adjustment is therefore required after assembly in order to keep the tuned speed within acceptable limits. Finally, the manufacturing technique is expensive and does not readily lend itself to large scale production. A study of patent applications shows a trend toward a semi-fabricated construction in recent years and suggests that the limitations of the 'carved-from-solid' approach have been recognised, though not all these ideas have been translated into production hardware.

The flexure hinge used in the Series 2000 design is fully fabricated from some 51 separate parts and is based upon an unusual 'pins-held' flexure pivot that has been used in our miniature gyroscopes for many years. Here, the philosophy was to fabricate the hinge from a large number of relatively simple components but, since many are identical, the final assembly contains only six different parts. The flexure hinge, Fig.6, takes the form of three concentric co-axial rings, interconnected via two pairs of cross spring flexure pivots to produce a universal joint. Each pivot is the size of a match head and is fabricated from twelve components. The pivots are buried within the wall thickness of the three rings, making a very compact assembly which is later cemented around the outside of the gas bearing. The two flexure blades in each pivot are photo-etched from precision rolled strip whose thickness is controlled to ± 60 micro inches (± 1.5 micron) and the relative position of the blades is controlled by parallel pins which fit into accurately machined holes in each of the two end flanges.

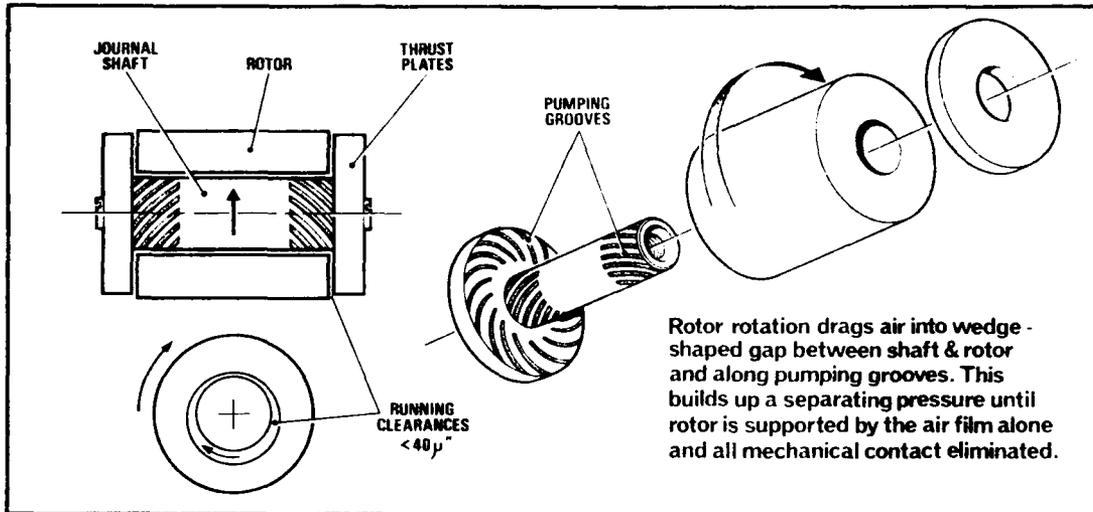


FIG 4 OPERATION OF H-FORM GAS BEARING

FIG 5 GAS BEARING COMPONENTS

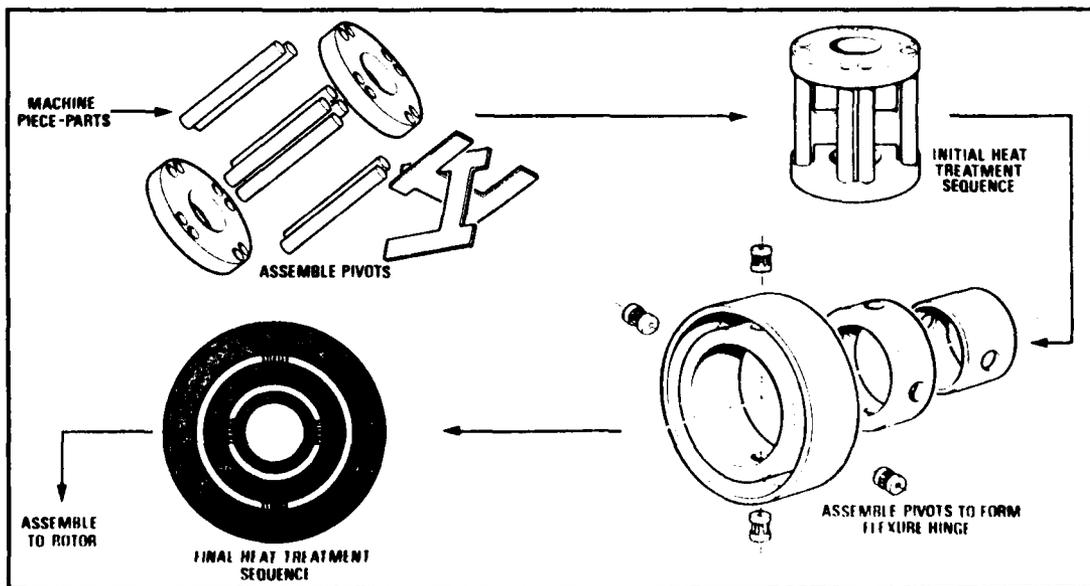


FIG 6 MANUFACTURING SEQUENCE FOR FLEXURE HINGE

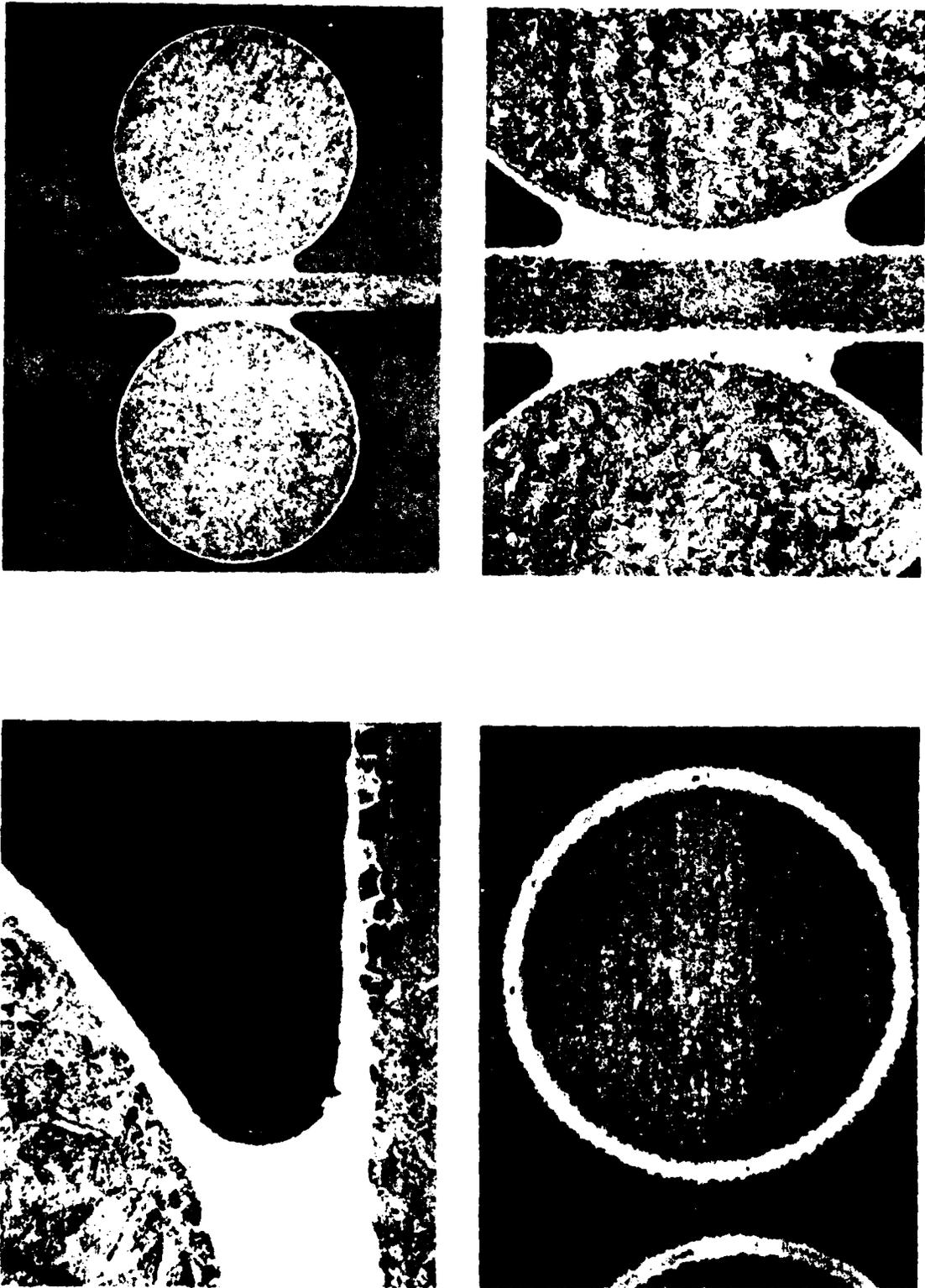


FIG 7

PHOTOMICROGRAPH SECTIONS
OF FLEXURE HINGE JOINTS

Fabrication is achieved by chemically depositing a very thin layer of a special alloy material onto certain components prior to assembly. The self jiggling assembly is then raised to a high temperature under vacuum where the alloy melts and is pulled into all the joint areas by capillary action. Surface tension forces in the liquid metal then centralise and align all the individual components and both blades, pins and pivot flanges float into their correct relative positions. Careful control of the process parameters allows limited penetration of the liquid braze into the substrate structure. This increases the strength of the joint and allows some of the base material to go into solution, thus raising the melting point of the braze and permitting successive brazing operations to be completed without disturbing the mechanical integrity of the original joints. Once all the joints have been established, further heat treatment sequences are used to condition the magnetic and mechanical properties of the structure and lastly, a controlled cooling sequence assures the final assembly is completely free from residual stress.

Designing the local geometry to take advantage of surface tension and capillary effects is no easy task but the benefits are considerable. The exact amount of braze material required to form each of the 104 joints in the hinge is automatically provided, with the surrounding surfaces acting as a braze reservoir. By using surface tension forces to align the individual components at high temperature, manufacturing and assembly stresses are virtually eliminated and the accuracy of the finished hinge is far higher than could ever be achieved by relying on the machining tolerances of the piece parts themselves. The local anchor point geometry of each flexure blade is defined by the shape of the final braze meniscus and can be closely controlled via the brazing temperature. [Where flexures are formed from solid or fabricated by conventional slot or abutment techniques, the anchor point geometry can vary considerably from part to part and cause corresponding variations in spring rate.] Because the flexure blade is formed from precision rolled strip prior to assembly, its physical characteristics can be assured with a very high degree of confidence and the grain flow orientation can be optimised for each blade. Similarly, the blade surface is highly polished and free from both surface and sub-surface metallurgical damage that could act as a stress-raiser and reduce its strength. Each blade is of constant section and does not need to be 'necked' to achieve the required spring rate. Each pivot is a true cross-spring device and it is known that cross spring suspensions have important performance advantages over other forms of suspension when applied to dynamically tuned sensors.

This unique design and manufacturing technique allows a very compact flexure hinge assembly to be produced, with a bore large enough to take a hydrodynamic gas bearing. The Figure of Merit (a dimensionless parameter indicating the 'goodness' of the design) is very high and the effect of mis-tuning errors correspondingly small. Finally, the method of manufacture readily lends itself to large scale production because many flexure assemblies can be processed simultaneously.

8. TORQUER DESIGN

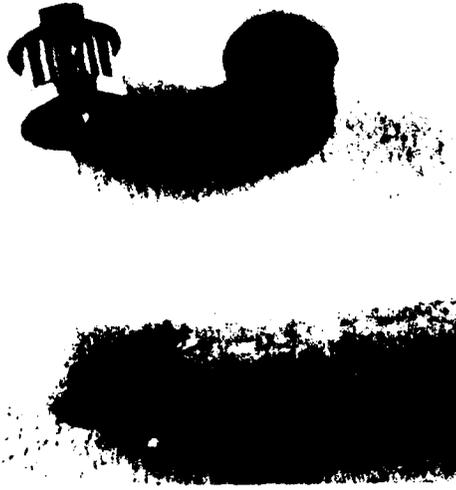
The torquer is a conventional D'Arsonval configuration, using a segmented Samarium Cobalt magnet and a high permeability return path. Encapsulated torquer coils are used to improve both the structural strength and the aerodynamic cleanliness of the design. Experimental work showed that capture rate would be limited by thermal distortion of the torquer coils rather than by electrical failure and the mounting and encapsulating arrangements are designed to reduce these effects to a minimum. Considerable effort was directed at optimising the magnetic design of the torquer to obtain the best possible capture rate consistent with the lowest practicable magnetic leakage. To this end, Finite Element analysis was used to predict the flux patterns within the magnetic circuit and the field strengths for comparison with those measured on experimental hardware. The Finite Element analysis was carried out on a large mainframe computer and the results subsequently used for the calculation of capture rate, inertias and masses. Optimisation of the torquer cannot be carried out in isolation and must of course be considered in conjunction with both gas bearing and flexure design.

9. THE CAPACITIVE PICK-OFF

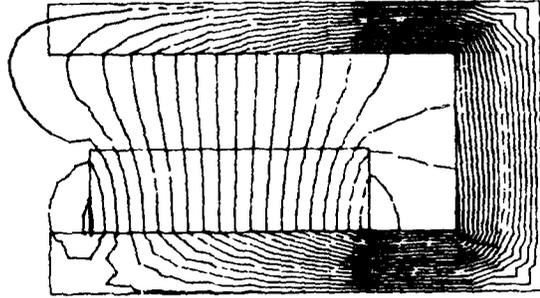
Inductive pick-offs are used extensively in existing strapdown DTGs. Their main disadvantage is that, however careful the design, they inevitably exert electro-magnetic torques upon the inertial element. An inductive pick-off was originally considered for the Series 2000 but it was soon apparent that the symmetrical design of the gyro readily lent itself to a capacitive version.

The resulting two-axis capacitive pick-off is mechanically simpler than the traditional inductive version and presents an aerodynamically 'clean' surface to the spinning inertial element. Windage forces acting on the inertial element are small and electrical forces are negligible.

The stationary pick-off plate is manufactured by a photo-etching process and is attached to one of the three modules adjacent to the torquer coils. Movement of the spinning inertial element about either input axis causes a change in air gap and this change is detected by one pair of capacitor plates positioned across the diameter of the wheel. Pick-off sensitivity varies inversely as the second power of the air gap length and the overall design of the sensor allows this mean gap dimension to be closely controlled. Mechanical contact between the stationary pick-off plate and the spinning surface is highly undesirable and a stop disc is therefore fitted on the opposite side of the inertial element to limit its angular travel to about 10 milliradians.



**FIG 8 FLEXURE PIVOT ASSEMBLY
PRIOR TO FINAL BUILD OF
FLEXURE HINGE**



**FIG 9 FINITE ELEMENT COMPUTER
PLOT OF TORQUER FIELD
DISTRIBUTION**

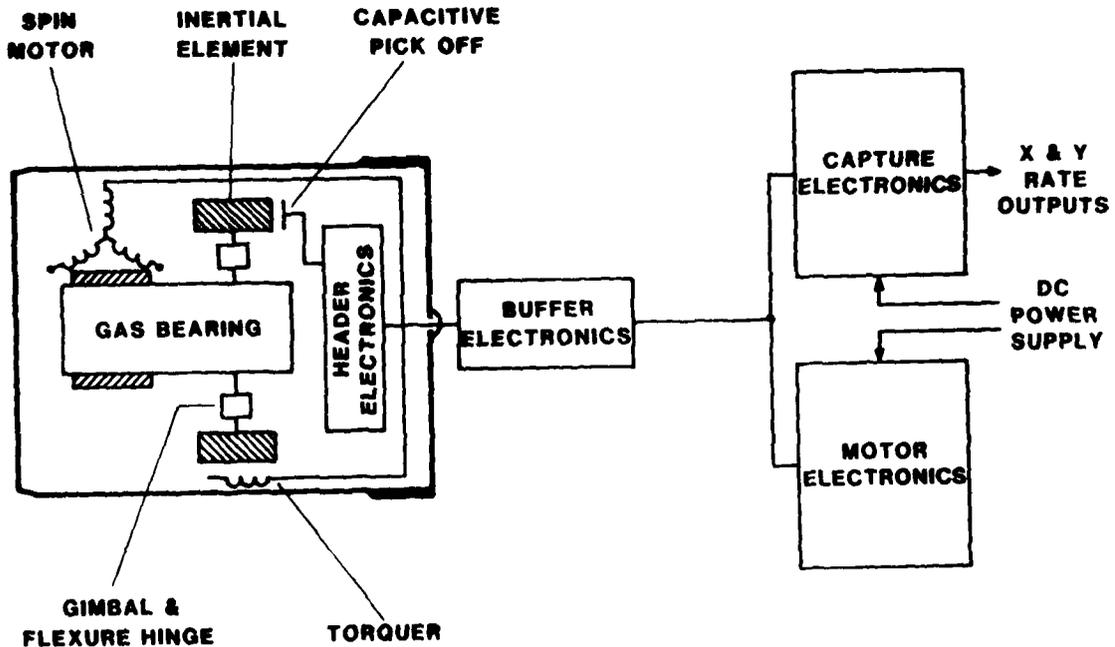


FIG 10 SCHEMATIC OF SERIES 2000 DTG AND ASSOCIATED ELECTRONICS

This stop disc spins with the flexure assembly to eliminate rubbing contact during overload conditions. The pick-off is energised from external electronics and de-coding circuitry is incorporated in header electronics which is sealed within the outer case.

10. ELECTRONICS

Supporting electronics was developed in parallel with the DTG hardware and is shown schematically in Figure 10. The purpose of the electronics is to generate a drive waveform to the spin motor, to energise and decode the capacitive pick-off and to provide the X and Y capture loops for the torquer. Spin motor drive is derived from a 4 MHz crystal oscillator and can be adjusted digitally in 0.1 Hz increments. The three-phase output to the motor windings comes directly from a totem pole configuration of FET power amplifiers and the drive frequency is compared with the derived reference via a phase locked loop and shift register system. Pick-off decoding is accomplished within the header electronics and pick-off energisation and bias voltage are generated within the buffer electronics which is sited close to the gyro. The X and Y capture loop circuits are identical and in the normal High Rate Loop mode, the torquer coils are current driven to assist in overcoming heating effects at high rates. Where Low Rate Loops are provided for test purposes, voltage is fed directly into the torquer coils because the heating effect is minimal.

The philosophy of the supporting electronics is illustrated in Figure 11. For test and evaluation purposes, the Basic Test Electronics is supplied in two small boxes complete with interconnecting cables and connectors. Both high and low rate loops are provided, together with ample test points and associated facilities. These modules contain plug-in circuit boards that can be exchanged for or supplemented by additional boards to modify the performance or characteristics of the system. Individual circuit cards are available separately to allow the customer to incorporate them into his own equipment and are referred to as the 'standard production electronics'. However, it is appreciated that for some applications, various 'non-standard' electronics will need to be developed to suit particular packaging and/or performance requirements and these are discussed with the customer concerned as appropriate. The current production electronics uses both CMOS and MOSFET technology. Hybrid power amplifiers are used in the capture loops but the remainder of the circuitry is discrete. Both the capture and motor circuitry are contained on small Eurocards and the buffer card measures 86 mm x 54 mm. Smaller discrete electronics have recently been developed to meet specific customer requirements and the intention is to fully hybridise both internal and external electronics in the near future. ± 15 V and ± 30 V. D.C. power inputs are standard.

To date, all performance requirements have been met without the phase locking, anti-hunting circuitry and various other refinements that are common to most contemporary DTGs, but the basic electronics described above have provision to incorporate such refinements at a later date if desired.

11. THERMAL DESIGN

Influenced perhaps by suggestions that certain 'first generation' DTGs dissipated 1 KW under maximum capture conditions, considerable attention was paid to the thermal design of the Series 2000. Choice of materials was dictated in many instances by the need to minimise thermal mis-match, and all adhesive joints include an expansion reservoir to assure the integrity of the joint under ambient temperature variations. Heat transfer across the gas bearing clearance is good and bearing dissipation can, for all practical purposes, be ignored. The Series 2000 configuration allows the spin motor stator to be in close thermal contact with the outer case and the very low quiescent power dissipation of the motor (< 1 watt) is therefore easily accommodated. Power dissipation in the header electronics is also negligible and this leaves the torquer as the major internal heat source.

Given that the magnetic design of the torquer is made as efficient as possible, the major problem is to extract the surplus heat from the torquer windings and direct it to the outside world via the shortest practical route. In the Series 2000, this is accomplished by bonding the four torquer coils to a solid copper ring, which is in turn thermally bonded to the outer case. This provides a very short heat path of high thermal conductivity and ensures low thermal gradients even under maximum torquing conditions. In terms of DTG performance, actual temperature is less significant than temperature gradient and the ability of the Series 2000 design to operate at relatively high case pressures is an advantage in this respect. The inertial element, gimbal and flexure hinges rely almost entirely upon windage effects to maintain them at an even temperature and the presence of a significant atmosphere within the case is therefore beneficial. As previously discussed, the cylindrical case of the DTG can be mounted with its entire length in contact with a heat sink, and this further minimises internal temperature gradients.

The combined effect of the above features has yielded a DTG with excellent thermal characteristics and with very low self heating in comparison with other designs. In terms of ambient temperature, the gas bearing frees the design from limitations imposed by the lubricants in conventional ball bearings and the Series 2000 will therefore operate over a wide ambient temperature range.

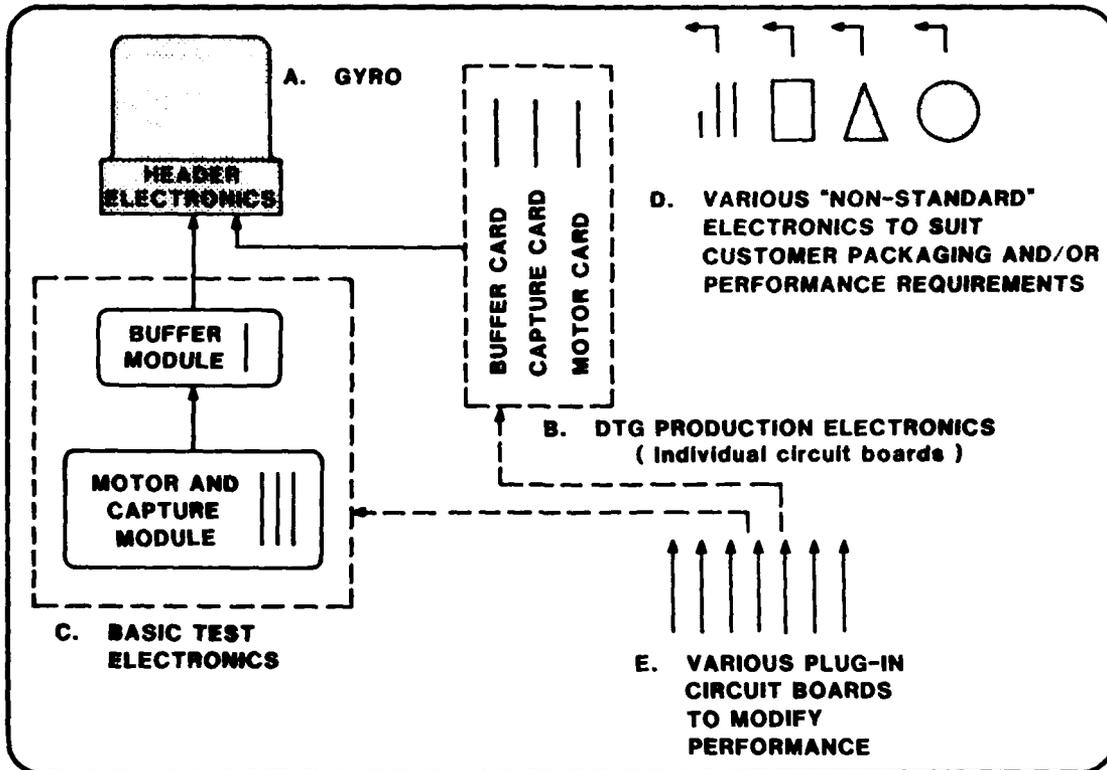


FIG 11 SUPPORT ELECTRONICS FOR THE SERIES 2000 DTG

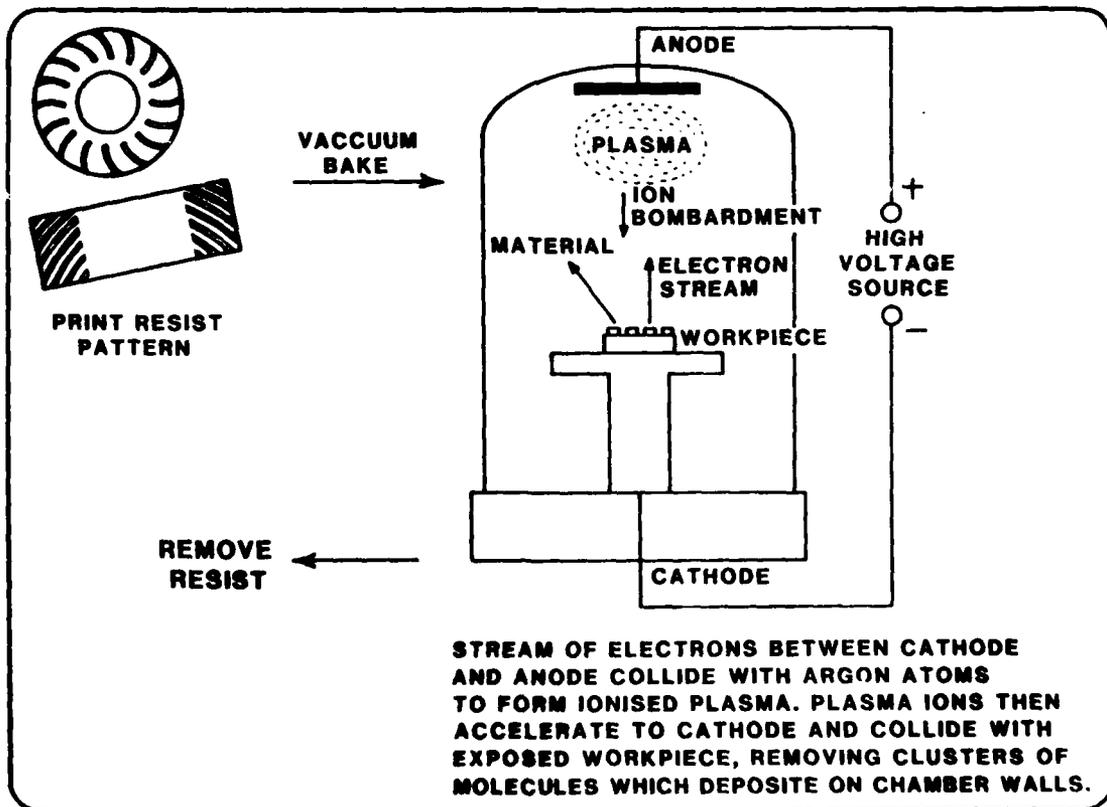


FIG 12 SCHEMATIC OF ION-MACHINING PROCESS USED TO CUT PUMPING GROOVES

12. MANUFACTURING TECHNOLOGY

Series 2000 manufacture involves a number of specially developed techniques and equipments. Gas bearing parts are machined from solid carbide blanks using high precision diamond grinding, lapping and polishing equipment. The final bearing surfaces are achieved by Thermochemical Polishing in which material is stripped from the surface by a thermal oxidation process to leave a very highly polished substrate that is smooth to a molecular level and free from normal polishing scratches. Pumping grooves are formed in the bearing surfaces by screen printing a special resist material onto the component and then cutting all the grooves simultaneously by Ion-Machining. Ion-Machining [4], may be likened to sand blasting, but using atomic particles in place of the sand and an electrical potential instead of high pressure air. It produces high definition grooves, typically 80 microinches (2 microns) deep whose depth can be controlled within close limits and which are free from all stress and machining debris. Ion-Machining is carried out in an argon plasma within a vacuum chamber and many components can be machined simultaneously at a typical machining rate of one micro inch per minute. Ion-Machining is an elegant process well suited to volume production and is considered to be the only practicable way of machining accurate pumping patterns in ultra hard materials such as Boron Carbide.

The manufacturing technique for the flexure hinge has been briefly described in Section 7 and relies initially on conventional manufacturing technology. However, special equipment had to be developed to position and size the holes for the flexure support pins and the final heat treatment sequences are carried out automatically in a computer controlled vacuum furnace. Assembly of the piece parts is, at the present time, a manual operation carried out under binocular microscopes, but the design is entirely self jiggling and no fixturing is necessary during heat treatment.

The manufacturing technique for the torquer coils is of interest because it allows a high packing density to be achieved and maximises the amount of copper within the torquer air gap. Two-dimensional coils are first wound on a low melting point disposable metal former and this is then rolled to form the familiar three-dimensional saddle-shaped arrangement prior to curing the impregnant and finally removing the metal support.

The specialised manufacturing facilities are supported by corresponding metrology equipment capable of verifying components machined to micro-inch tolerances. As for the manufacturing plant, much of this metrology equipment was either developed or adapted in-house. Considerable experience in machining exotic materials was already available within the company but many new skills had to be learnt to cope with the wide range of different materials used throughout the Series 2000 DTG.

13. ASSEMBLY AND TESTING

The modular design of the Series 2000 DTG is a considerable advantage during the build stage, when a suspect module can be quickly replaced prior to rectification. Each of the three major modules are built and tested separately prior to fitting together for the first time. Preliminary functional testing of this three-part assembly can be carried out prior to fitting the outer case, which is initially sealed with an 'O' ring. Final hermetic sealing is completed at a later stage once provisional performance data has been obtained. A comprehensive sequence of burn-in and testing is carried out throughout all stages of build. Multi-position and rate table testing are used to establish the quality of the finished sensor and much of this data is obtained and recorded automatically via computer driven data logging equipment. Gas bearing assembly must be carried out under stringent Class 100 conditions involving specialised cleaning and handling techniques but once the module containing the gas bearing is assembled, all subsequent work can be completed in a normal gyro clean room environment. Cements and adhesives are used extensively throughout the sub-assembly sequence and post assembly cleaning includes vacuum baking to remove all traces of volatile contaminants that could degrade bearing performance. Although the use of screw threads has been reduced to an absolute minimum in this design, it is nevertheless a fully repairable item.

14. ADVANTAGES OF THE SERIES 2000

The advantages claimed for DTGs in general are listed in Figure 14 and the major features of the Series 2000 design are summarised in Figure 15. They are largely self explanatory but some comment is in order.

Frequent claims have been made that DTG performance is largely independent of spin bearing quality because the inertial elements are effectively de-coupled from the bearing and drive system by the flexure hinge. However true this may be in theory, it is certainly not true in practice. Our own experience confirms that performance is not independent of the spin bearings and that a very significant improvement is obtained when ball bearings are replaced by a gas bearing. The gas bearing has proved easier to assemble than corresponding ball bearings (which require rather critical pre-loading and alignment arrangements) and has the following advantages over the latter, viz:



**FIG 13 SUB ASSEMBLY MODULES
PRIOR TO FINAL ASSEMBLY**

DYNAMICALLY TUNED GYROSCOPES

- TWO AXIS RATE SENSING
- DRY CONSTRUCTION - NO LIQUIDS
- I.N. POTENTIAL AT RELATIVELY LOW COST
- STRAPDOWN OR PLATFORM CAPABILITY
- SMALL COMPACT SIZE

**FIG 14 ADVANTAGES OF
DYNAMICALLY TUNED GYROS**

SERIES 2000 DTG

- UNIQUE CONFIGURATION - NO OVERHANG
- GAS BEARING
- FABRICATED FLEXURE ASSEMBLY
- CAPACITIVE PICK-OFF
- SIMPLE MODULAR CONSTRUCTION
- IMPROVED HEAT DISSIPATION
- HIGH EFFICIENCY SPIN MOTOR
- INTEGRAL HEADER ELECTRONICS
- VERY HIGH FIGURE OF MERIT
- PROVEN SI TECHNOLOGY
- INDIGENOUS SI DESIGN

**FIG 15 FEATURES OF THE
SERIES 2000 DESIGN**

SERIES 2000 PERFORMANCE

- CAPTURE RATE

| | |
|--------------|-----------|
| Continuous | 200 deg/s |
| Intermittant | 420 deg/s |
- DRIFT

| | |
|---------------|-----------------|
| Random-in-run | 0.1 deg/h (10') |
| Run-to-Run | 1 deg/h (10') |
| G-insensitive | 10 deg/h |
| G-sensitive | 10 deg/h |
- NON-LINEARITY < 0.01 deg/s
- HYSTERESIS << 0.01 deg/s
- SCALE FACTOR REPEATABILITY 0.01%
- RUN-UP TIME 2 1/2 s
- QUIESCENT POWER < 1 watt
- DYNAMIC RANGE 3.3 x 10'

**FIG 16 TYPICAL PERFORMANCE
ACHIEVED TO DATE**

- * Unlimited Storage Life.
- * Very high operational life and MTBF
- * Very low acoustic noise
- * Rapid run-up over a wide temperature range
- * Highly stable and repeatable spin axis definition

15. TYPICAL PERFORMANCE

Typical performance currently being achieved from development hardware in tests at RAE and at Smiths Industries is summarised in Figure 16. The performance compares extremely favourably with that of comparable tuned sensors in this class and this combination of drift performance and capture rate is attractive over a wide range of applications. This level of performance has been achieved, as previously stated, with relatively basic electronics and after a fairly short development time and is clearly capable of further improvement.

The current design standard is maturing rapidly and unit-to-unit performance variations are small. Each gyro has an internal temperature sensor that is user accessible, allowing external compensation or characterisation to be applied as required. Extensive testing to date has shown performance variation with temperature to be exceptionally repeatable and predictable, thus allowing accurate modelling of these sensor characteristics within a typical processor based system. In-run drift performance depends to some extent upon case pressure and can be traded against other parameters as previously discussed.

16. APPLICATIONS

DTGs in general are already being specified for a wide range of applications ranging from inertial navigation to missile control and the unique attributes of the Series 2000 design are expected to further broaden this range of opportunities. Figure 17 details the range of applications foreseen in the immediate future, many of which are under active discussion with potential users. The Series 2000 DTG is currently being evaluated in a medium grade inertial navigation system and in this type of application, the high MTBF available from the gas bearing is usually an important consideration. For applications involving missiles, munition dispensers and sub-munitions, long storage life is becoming increasingly important and here again, the inherent characteristics of the self generating gas bearing make it the obvious choice for this role. Similarly, rapid run-up time across a wide range of temperatures can often be a deciding factor in the choice of sensors for tactical weapon systems and once again, gas bearing technology has few equals in this respect. In applications involving underwater weapons and systems, the very low acoustic noise generated by a hydrodynamic gas bearing (approximately 1/50 of the mean spectral noise density of a high quality ball bearing) can be a deciding factor.

The current emphasis on cost effective guidance and control systems places particular constraints on the primary sensors, especially in terms of affordability. Sensor performance may be superb, but if it is achieved by lengthy fiddling and tweaking and is consequently unaffordable, the system designer will be obliged to select a lesser device or to seek a solution elsewhere. The Series 2000 concept is an attempt to reduce the need for these protracted adjustments by applying proven state-of-the-art technology at the design and manufacturing stages, so increasing the price/performance ratio of the final product. Life and reliability have been the traditional 'Achilles Heel' of rotating mass sensors and have driven the continuing search for potentially reliable alternative devices. The gas bearing effectively removes this limitation from the tuned sensor and allows it to compete on more than equal terms with contemporary optical and solid state rate sensors whose size, cost and complexity are frequently unfavourable by comparison.

17. CURRENT STATUS

Research and development of the Series 2000 DTG is complete and the sensor has now entered production at Cheltenham. The first production version is designated the 2001 DTG and will form part of a family of DTGs intended for both general and specific applications. Pilot production of customer evaluation hardware commenced towards the end of 1983 and performance testing of prototype gyros continues in order to consolidate the data base and identify areas of future development.

18. FUTURE POTENTIAL

It is clear that the basic mechanical and electronic design has potential for improvement. Drift improvement down to a true IN level should be achievable in the foreseeable future and certain other parameters can be tailored to meet specific applications as required. It is envisaged that these improvements will be achieved by relatively small modifications to the existing hardware and building techniques rather than by changes to the basic design. A micro-miniature (less than one inch diameter) version is already under consideration and this would be capable of capturing input rates in excess of 1000 deg/s and would have an energisation time of less than one

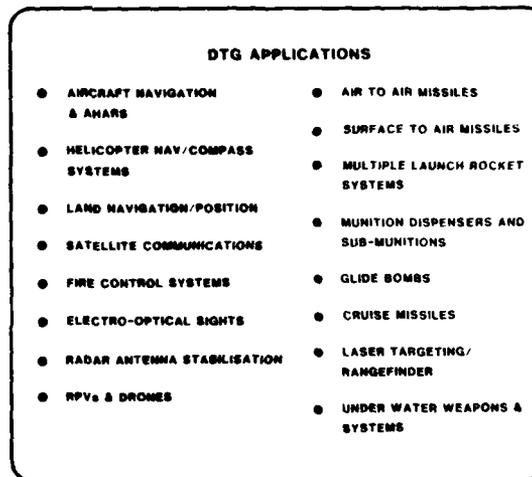


FIG 17 POTENTIAL APPLICATIONS FOR THE FOR THE SERIES 2000 DTG

second. Quite apart from the obvious advantages of size, weight, capture rate and ready time, the dimensions and design of this sensor will ensure it is rugged enough to cope with the most extreme environmental conditions, as typified by vertical launch and vectored thrust missile systems.

REFERENCES

1. BEARDMORE G., 'A Review of Current Gas Bearing and Flexure Pivot Technology at Smiths Industries', Joint MOD/US Fluid Film Bearing Meeting, Southampton Univ., 1982.
2. BEARDMORE G., 'Development of the Series 700 Gas Bearing Gyro', Proc. 5th Internat. Gas Brg. Symp. Southampton Univ., Paper 7, 1971.
3. BEARDMORE G., 'The Steady-State Performance of an H-Form Aerodynamic Gyro Bearing Having Connected Journal and Thrust Spiral Grooved Surfaces; With Particular Reference to its Operational Life and Manufacturing Limitations', Ph.D Thesis, University of Aston, September 1974.
4. BEARDMORE G., 'Ion-Machining, a Technique for Cutting Pumping Grooves in Gas Bearing Components', Proc. 5th Internat. Gas Brg. Symp., Southampton Univ., Paper 8, 1971.

ACKNOWLEDGEMENTS

The author would like to thank the directors of Smiths Industries Aerospace & Defence Systems - Cheltenham - for their permission to publish this paper. He would also like to acknowledge the technical support received from staff in the Radio and Navigation Department at Farnborough throughout the entire research and development programme. Computational assistance was also received from Dr. D. Howe at the Electrical Engineering department of the University of Sheffield and Dr. C.H.J. Fox of The University of Nottingham acted as analytical consultant during the initial phases of the project. Finally, the author must pay tribute to all those who worked on the project at Cheltenham and who made the Series 2000 DTG a reality.

SYSTEME DE GUIDAGE A GYROLASER POUR MISSILES A MOYENNE PORTEE
GYROLASER GUIDANCE SYSTEM FOR MEDIUM RANGE MISSILES

par
Bernard de Salaberry
Chef du Département Inertie et Optique
Société Française d'Équipements pour la Navigation Aérienne (SFENA)
Aérodrome de Villacoublay
78141 VELIZY-VILLACOUBLAY - FRANCE

RESUME

La technologie du gyrolaser a maintenant évolué de sorte que ce capteur est devenu utilisable sur des missiles tactiques de moyenne portée. Après une présentation des performances requises pour les capteurs, on trouvera une description des capteurs, gyrolasers et accéléromètres, retenus pour réaliser un ensemble inertiel destiné au guidage et au pilotage de tel missile. Les problèmes de filtrage du bruit sur les signaux de lecture des gyrolasers sont examinés.

ABSTRACT

Laser gyro technology is now developed to the point where this type of inertial sensor can be used on medium-range tactical missiles. After outlining the performances required for these sensors, we shall go on to describe the laser gyros and accelerometers chosen to construct an inertial system designed for the guidance and control of such a missile. The problems of filtering noise on the laser gyro readout signals will also be examined.

1. INTRODUCTION

Au cours des vingt dernières années, les missiles tactiques ont subi une évolution profonde et une grande diversification, tant dans leur taille que dans leur portée, leur mission et leurs performances.

Du moment qu'ils étaient guidés, ils étaient équipés de senseurs inertiels fournissant des références pour leur guidage. Ces senseurs étaient plus ou moins sophistiqués : depuis les gyroscopes à poudres pour les missiles antichar jusqu'aux centrales à composants liés des missiles les plus modernes en passant par des centrales à deux gyroscopes, des plateformes inertielles pseudo-strapdown à la fin des années 60 et des plateformes trois axes plus récemment.

Les projets de missiles tactiques actuels, dès lors que leur portée nécessite un guidage inertiel même peu précis, prévoient tous l'emploi de centrales à composants liés.

Pour sa part, la SFENA, envisageant une telle évolution, avait dès le début des années 1970 recherché quel serait le capteur gyrométrique le mieux adapté à une utilisation à composant lié dans un grand domaine de mesure et dans un environnement sévère. Son choix s'est porté sur le gyrolaser car il apparaissait clairement que ce capteur pouvait, de par sa conception, avoir de très bonne performance en dérive sans limitation de domaine de mesure autre que celle que pourrait apporter l'électronique utilisée pour traiter les signaux de sortie.

L'emploi du gyrolaser a pu paraître trop luxueux pour des applications aux missiles tactiques et c'est l'une des raisons pour laquelle la plupart des industriels spécialistes du gyrolaser ont jusqu'ici privilégié le développement de gyrolasers très performants applicables à la navigation pour avions.

Il semble aujourd'hui que les performances demandées aux systèmes inertiels des missiles tactiques futurs rentrent tout-à-fait dans la gamme de performances réalisables avec des gyrolasers de petite taille.

La technologie du gyrolaser a par ailleurs évolué de telle sorte que des petits gyrolasers sont aujourd'hui réalisables industriellement et à des coûts qui les rendent parfaitement compétitifs vis-à-vis des gyromètres classiques.

La SFENA est dès aujourd'hui en mesure de fournir des systèmes inertiels équipés de gyrolasers et dont l'encombrement, le prix et la facilité de maintenance sont bien adaptés à un emploi sur des missiles de moyenne portée tels des missiles anti-navires.

De tels systèmes ont généralement la configuration de la figure 1.

Trois gyromètres et trois accéléromètres fournissent des signaux qui ne peuvent être utilisés directement pour la navigation et le pilotage, soit parce qu'il faut effectuer des corrections de biais et de facteur d'échelle, soit parce qu'ils sont trop bruités pour être utilisés directement pour le pilotage ou la stabilisation de la ligne de visée de l'auto-directeur.

Les calculs nécessaires à la navigation ont été très souvent décrits. Ceux qui concernent le pilotage seront spécifiques au missile. C'est pourquoi, cet exposé ne traitera que des aspects liés aux capteurs et à l'élaboration des signaux nécessaires à la navigation et au pilotage.

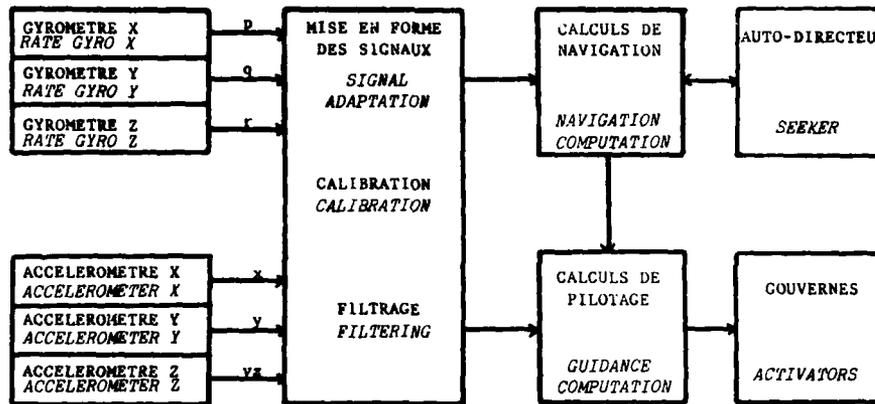


FIGURE 1
SYNOPTIQUE DU GUIDAGE INERTIEL D'UN MISSILE
INERTIAL GUIDANCE FOR MISSILE - TYPICAL ORGANIZATION

La suite de cet exposé traitera :

- des spécifications des capteurs pour les utilisations sur missile tactique de moyenne portée,
- des choix technologiques concernant le système, les gyrolasers et les accéléromètres,
- des problèmes de bruit sur les signaux destinés au pilotage.

2. SPECIFICATIONS DES CAPTEURS

Puisque les missiles moyenne portée sont généralement des missiles à préguidage inertiel, la précision des capteurs va dépendre des performances demandées au préguidage.

Lorsque l'on parle de moyenne portée, il s'agit généralement de portées comprises entre 30 et 300 km, ce qui est très étendu.

La précision demandée au préguidage sera telle que l'erreur de navigation soit très inférieure aux incertitudes sur la position de la cible au moment où l'auto-directeur est en mesure de la repérer. Ces incertitudes dépendent évidemment de la distance de la cible mais aussi de sa vitesse et de la vitesse du missile.

L'éventail des hypothèses possibles est donc considérable ; cependant, en prenant des cas de trajectoires sur les plus longues distances, on peut fixer des performances nécessaires pour l'inertie et en déduire des spécifications pour les capteurs.

Pour ce qui concerne le pilotage, les caractéristiques à spécifier prendront en compte le type de trajectoire, les modes de pilotage et les précisions recherchées à l'impact.

Le tableau de la figure 2 ci-dessous donne un ordre de grandeur des principales spécifications pour les gyromètres et les accéléromètres :

| | GYROMETRE | ACCELEROMETRE |
|---|---------------------------------------|---|
| Stabilité du biais <i>Drift</i> | 0,5 à 20°/h | 2.10 ⁻⁴ g à 2.10 ⁻³ g |
| Marche au hasard <i>Random walk error</i> | 0,1 à 1°/h | |
| Stabilité du facteur d'échelle <i>Scale factor stability</i> | 5.10 ⁻⁵ à 10 ⁻³ | 10 ⁻⁴ à 10 ⁻³ |
| Linéarité du F.E. <i>Scale factor linearity</i> | 5.10 ⁻⁵ à 10 ⁻³ | 10 ⁻⁴ à 10 ⁻³ |
| Domaine de mesure <i>Range</i> | 400 à >2000°/sec | 200 à 1000 m/sec ² |
| Résolution <i>Sensitivity</i> | 1 à 10 sec. arc/pulse | 0,5 à 5 cm/sec |

FIGURE 2
PRINCIPALES SPECIFICATIONS DES GYROMETRES ET DES ACCELEROMETRES

On peut remarquer que les domaines de mesure tendront toujours à augmenter aussi bien pour les vitesses angulaires que pour les accélérations. Si l'on veut qu'un système puisse être utilisé aisément sur plusieurs types de missiles, il faut qu'il soit capable des plus grands domaines de mesure.

Les autres spécifications vont concerner le volume, le poids, la consommation, le domaine de température et vont dépendre étroitement du type de missile et de sa mission. Le meilleur compromis doit être trouvé entre la miniaturisation, la fiabilité, la facilité de maintenance et les coûts. Enfin, les capteurs doivent être prévus pour une durée de vie en stockage supérieure à 15 ans et permettre d'espacer le plus possible les contrôles périodiques pour minimiser les coûts d'entretien des missiles.

3. CHOIX DES CAPTEURS

3.1. Le gyromètre

Dès lors que l'on recherche des stabilités de biais de quelques degrés par heure dans un domaine de mesure pouvant aller au-delà de 2000°/sec., le gyromètre laser est le capteur le mieux adapté. En effet, ces deux caractéristiques, stabilité du biais et domaine de mesure ne dépendent pas des mêmes paramètres physiques. L'un dépend de la qualité de l'optique, l'autre de la bande passante de l'électronique. Et l'on peut aisément améliorer l'un sans détériorer l'autre, ce qui n'est pas le cas avec les gyromètres mécaniques classiques.

Avant de développer un gyromètre laser pour missile, il faut faire certains choix technologiques.

3.1.1. Choix de la forme et de la taille

Compte tenu des performances demandées et des contraintes de volume, la taille sera généralement petite et le choix est possible entre une forme triangulaire et une forme carrée pour le bloc optique.

La figure 3 montre que pour les petites tailles, la forme carrée ne présente pas d'avantage évident pour l'encombrement par rapport à la forme triangulaire (à sensibilité identique bien entendu ; dans ce cas, un carré de 9,2 cm est équivalent à un triangle de 12 cm).

En outre, on voit qu'il n'y a plus de place sur le côté pour placer la cathode et qu'il faudra donc la placer sur le dessus, ce qui sera un inconvénient pour la fixation du bloc optique sur un support et pour l'encombrement du bloc lui-même.

Au plan des coûts et après industrialisation, un gyrolaser triangulaire sera moins onéreux puisqu'il ne compte que trois miroirs et que les miroirs sont un point important du prix du gyrolaser.

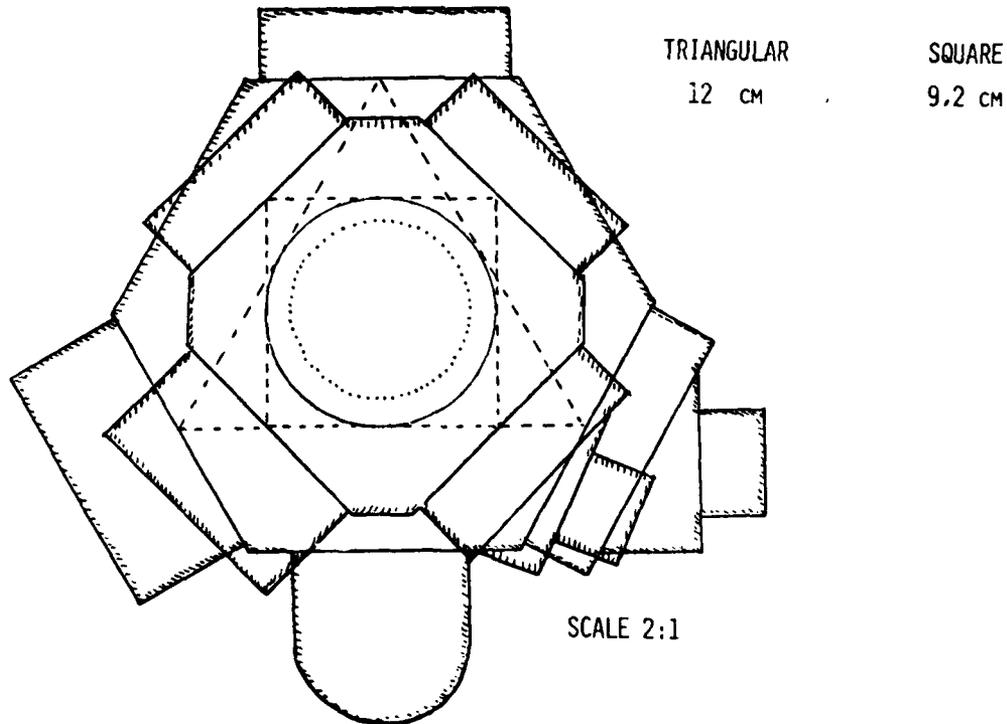


FIGURE 3
COMPARAISON DES FORMES TRIANGULAIRES OU CARRÉES
TRIANGULAR OR SQUARE SHAPE COMPARISON

Pour le choix définitif du périmètre du gyrolaser, il faut tenir compte de l'ensemble des performances demandées. La technologie des miroirs aujourd'hui progresse de sorte qu'il est possible de réaliser des gyrolasers avec des périmètres aussi petits que 6 cm.

Le tableau de la figure 4 ci-dessous montre les niveaux de performances qui peuvent être obtenus pour différents périmètres de gyrolaser triangulaire.

| Périmètre cm | 33 | 21 | 12 | 6 |
|---|-------------------------------|---------------------------------------|---------------------------------------|-----------------------|
| Stabilité du biais <i>Drift °/h</i> | 0,003 à 0,03 | 0,01 à 0,1 | 0,1 à 1 | 0,5 à 5 |
| Marche au hasard <i>Random walk °/√h</i> | $2 \cdot 10^{-3}$ à 10^{-2} | $6 \cdot 10^{-3}$ à $3 \cdot 10^{-2}$ | $3 \cdot 10^{-2}$ à 0,15 | 0,2 à 1 |
| Facteur d'échelle <i>Scale factor pulse/°/sec</i> | 1800 | 1145 | 655 | 327 |
| Stabilité du facteur d'échelle <i>Scale factor stability</i> | 10^{-6} à 10^{-5} | $2 \cdot 10^{-6}$ à $2 \cdot 10^{-5}$ | $3 \cdot 10^{-6}$ à $3 \cdot 10^{-5}$ | 10^{-5} à 10^{-4} |

FIGURE 4
GAMME DE PERFORMANCES DES GYROLASERS EN FONCTION DU PERIMETRE

Pour chaque performance, les valeurs les moins bonnes correspondent à ce qui peut être obtenu sans précaution particulière et donc au moindre coût, les valeurs les meilleures sont une indication des améliorations qui peuvent être apportées sur chaque type de gyrolaser soit dans sa définition, soit par une modélisation plus poussée des erreurs.

On voit aisément qu'un gyrolaser de 12 cm de périmètre permet de tenir facilement toutes les performances. Un gyrolaser de 6 cm peut également être satisfaisant mais il risque de coûter plus cher à performance égale puisqu'il faudra lui apporter des améliorations.

Deux éléments seront également déterminants pour le choix. Il s'agit de la stabilité des performances pendant le stockage et du niveau de bruit pour le pilotage.

Pour le premier point, des performances supérieures à celles requises vont permettre d'espacer, voire de supprimer les contrôles pendant la vie du missile. Dans ce cas, le gyrolaser de 12 cm est beaucoup plus intéressant que celui de 6 cm.

Pour le second point, le bruit de quantification est évidemment multiplié par deux en passant de 12 à 6 cm. Le bruit de marche au hasard est lui multiplié par un facteur beaucoup plus important dans des zones de fréquence où le pilotage sera sensible (figure 5). Ce point particulier peut dans certains cas être dimensionnant et être une limitation pour l'emploi de très petits gyrolasers.

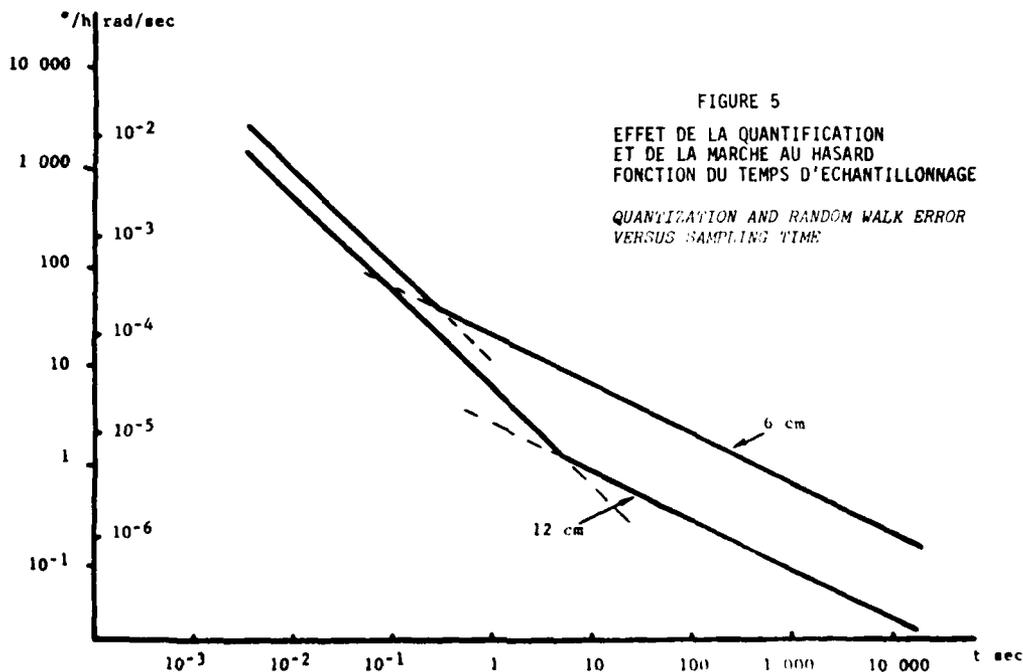


FIGURE 5
EFFET DE LA QUANTIFICATION
ET DE LA MARCHÉ AU HASARD
FONCTION DU TEMPS D'ECHANTILLONNAGE
QUANTIZATION AND RANDOM WALK ERROR
VERSUS SAMPLING TIME

Les critères d'encombrement imposés pour les missiles tactiques de moyenne portée pendant la prochaine décennie sont tels qu'ils n'imposent pas l'emploi d'un très petit périmètre. C'est pourquoi la SFENA a décidé de développer des gyrolasers de 12 cm pour réaliser des systèmes inertiels pour ce type de missiles.

3.1.2. Dispositif d'élimination de la zone aveugle

Les gyrolasers présentant une zone aveugle due au couplage des ondes lumineuses sous l'effet des rétrodiffusions des miroirs, il faut les munir d'un dispositif capable d'éliminer cette zone aveugle.

Le dispositif d'élimination de la zone aveugle retenu malgré ses inconvénients est l'activation mécanique qui consiste à faire osciller le bloc optique sur son axe de mesure à une fréquence de quelques centaines de hertz et avec une vitesse crête comprise entre 100 et 200°/sec.

C'est encore aujourd'hui le seul dispositif qui ne détériore pas les performances des gyrolasers. Il utilise l'inertie du bloc optique monté sur un système élastique en rotation. L'ensemble oscille à sa fréquence propre sous l'action d'un moteur commandé par des circuits électroniques adaptés.

Pour des raisons pratiques, il s'est avéré très difficile de l'incorporer au bloc optique et il est placé sous celui-ci. Pour éviter tout mouvement conique important, les trois fréquences d'activation sont volontairement décalées les unes par rapport aux autres. L'ensemble est dimensionné pour tenir les conditions d'environnement mécanique les plus sévères (figure 6).

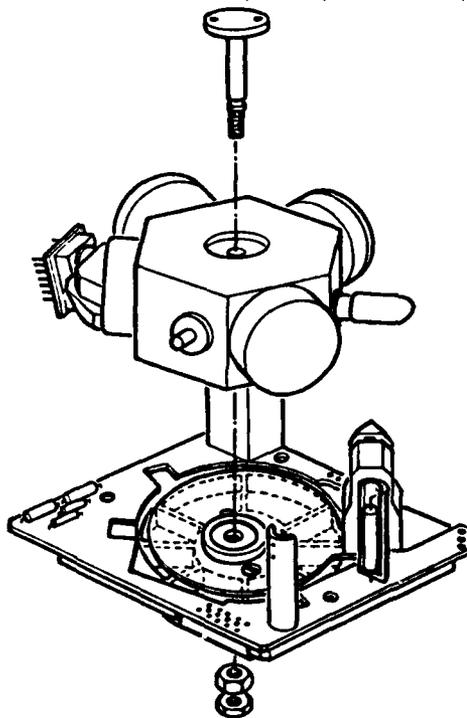


FIGURE 6
SOUS-ENSEMBLE GYROLASER
LASER GYRO SUB-ASSEMBLY

3.1.3. Dispositif de lecture

Il existe une solution optique pour éliminer le bruit sur la vitesse angulaire mesuré par le gyrolaser dû au mouvement alternatif d'activation. Cette solution consiste à fixer deux éléments du système de lecture sur le boîtier et non sur le bloc optique et à utiliser le mouvement relatif du bloc et du boîtier pour créer un défilement de frange égal et opposé à celui dû à l'effet gyrométrique lui-même. Le déplacement des franges dû à l'activation est ainsi annulé et il ne subsiste plus que le défilement utile.

Ce système est assez délicat à réaliser. Le filtrage électronique des signaux d'activation a paru à terme moins coûteux et a été retenu. C'est pourquoi, le dispositif de lecture est classique et fixé directement sur le bloc optique.

Les autres circuits électroniques associés au gyrolaser et qui comprennent l'asservissement de longueur de cavité, la régulation du courant et le dispositif d'allumage sont classiques.

L'ensemble des circuits électronique est rassemblé dans trois circuits hybrides couches épaisses. Ils seront ultérieurement réalisés en circuit prédiffusé, ce qui permettra de réduire les coûts.

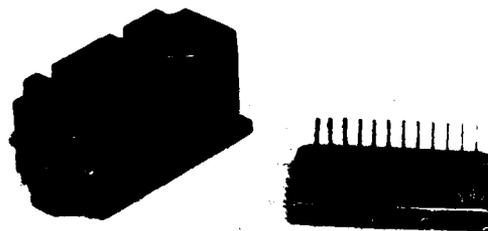


FIGURE 8
ACCELEROMETRE J125
J125 ACCELEROMETER

Pour réduire l'encombrement, les trois gyrolasers sont dépourvus de boîtiers. Ils sont logés dans les alvéoles d'une pièce en fonderie qui constitue l'unité de mesure inertielle (UMI) du système et assure une très grande rigidité et une bonne stabilité entre les axes de mesures (figure 7).

Le volume de l'unité de mesure inertielle seule ne dépasse pas 1,5 l et le boîtier complet avec le calculateur et les alimentations a un volume de 3,3 l.

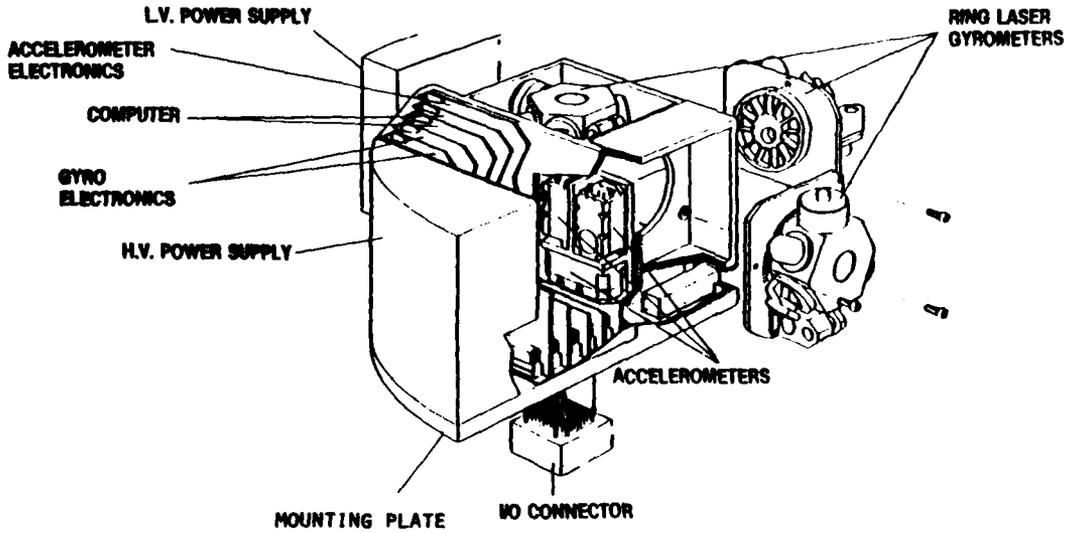


FIGURE 7
SYSTEME INERTIEL POUR MISSILE - ARCHITECTURE
INERTIAL UNIT FOR MISSILE - ORGANIZATION

3.2. L'accéléromètre

Les conditions de choix pour l'accéléromètre sont :

- Un domaine de mesure important ;
- Une bonne stabilité du biais ;
- Une grande résistance aux conditions d'environnement mécanique et notamment aux chocs violents contre lesquels ils ne seront pas protégés par une suspension.

La SFENA dispose d'une grande gamme d'accéléromètres pendulaires asservis.

Parmi ceux-ci, le modèle J 125 (figure 8) dispose d'une suspension à pivots très robuste. Il est réalisé dans une technologie faible coût qui répond très bien à ce besoin et son domaine de mesure peut être ajusté en fonction du besoin.

Ses principales caractéristiques sont représentées sur la figure 9 ci dessous :

Son asservissement est réalisé en binaire forcé, ce qui permet d'obtenir directement des incréments de vitesse et évite une conversion tension fréquence toujours onéreuse.

| | | | | |
|--|---------------------|----------------------|----------------------|----------------------|
| Domaine de mesure Range | m/s ² | 65 | 250 | 500 |
| Facteur d'échelle Scale factor | Hz/m/s ² | 10 ³ | 256 | 128 |
| Stabilité du facteur d'échelle Scale factor stability | | 10 ⁻⁴ | 10 ⁻⁴ | 2.10 ⁻⁴ |
| Variation du F.E. en tempér. S.F. Variation with temp. | /°C | 1.4 10 ⁻⁴ | 1.4 10 ⁻⁴ | 1.4 10 ⁻⁴ |
| Stabilité du biais Bias stability | m/s ² | 10 ⁻³ | 10 ⁻³ | 2.10 ⁻³ |
| Variation du biais en tempér. Bias variation with temp. | m/s ² | 8.10 ⁻⁴ | 3.10 ⁻³ | 6.10 ⁻³ |

FIGURE 9 : ACCELEROMETRE J125 - PRINCIPALES CARACTERISTIQUES

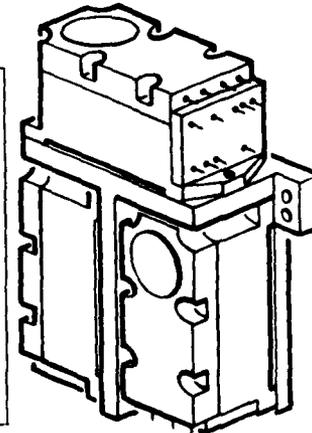


FIGURE 10
TRIDRE ACCELEROMETRIQUE
TRI-AXIAL ACCELEROMETER SUB-ASSEMBLY

Les trois accéléromètres sont placés sur un trièdre support fixé lui-même sur l'UMI (figure 10).

4. TRAITEMENT DES SIGNAUX DE PILOTAGE

Les critères de choix pour le traitement des signaux de sortie pour le pilotage sont les suivants :

- Retard de datation faible ;
- Fréquence de sortie pour le pilotage assez élevée (300 à 500 Hz)
- Spectre de bruit aussi réduit que possible.

Il est évident que les deux premiers critères sont liés. Mais le retard de datation va aussi dépendre de la fréquence d'activation des gyrolasers et du procédé utilisé pour filtrer le mouvement d'activation.

4.1. Filtrage des mouvements d'activation des gyrolasers

Trois procédés sont possibles :

Le premier (figure 11), équivalent au filtrage optique, consiste à mesurer la vitesse du bloc optique par rapport au boîtier, ou dans notre cas au trièdre de l'unité de Mesure Inertielle. Cette mesure, convertie en fréquence, est ensuite soustraite de la mesure issue du gyromètre à l'aide d'un compteur décompteur.

Ainsi, le mouvement dû à l'activation est compensé et si les réglages sont bien faits, il ne subsiste plus de bruit d'activation à la sortie du compteur décompteur (figure 13).

Le second et le troisième procédé (figure 12) effectuent le filtrage par calcul, par addition des impulsions acquises pendant une période d'échantillonnage et de celles acquises pendant la période précédente. On voit aisément que si l'acquisition est faite en synchronisme avec l'activation mécanique du gyrolaser, c'est-à-dire si $F_c = F_d$ ou si $F_c = 2F_d$, ce qui est plus favorable, les impulsions de sorties du gyromètre dues au mouvement d'activation vont s'éliminer.

Dans le cas du second procédé, le calcul est fait par logiciel dans le calculateur.

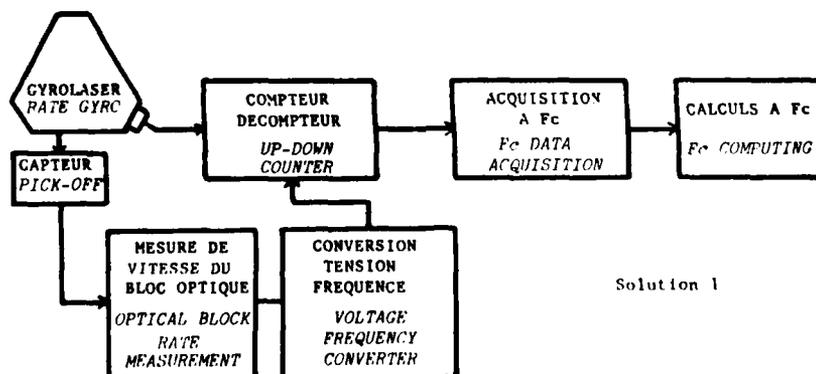
Malheureusement, du fait que pour éviter des mouvements coniques, les fréquences d'activation des trois gyrolasers sont légèrement différentes, le prélèvement synchrone par le calculateur est quasiment impossible sans risquer de perdre des informations ou sans rajouter des bruits supplémentaires. Le prélèvement sera donc asynchrone, à une fréquence voisine du double de la fréquence d'activation. La compensation d'un prélèvement par le prélèvement précédent est imparfaite et entraînera un bruit résiduel dont la fréquence principale sera $2F_c \approx F_d$ dont l'amplitude crête sera $\sim d \times 2\pi (2F_c - F_d)$.

Bien qu'étant hors du domaine des fréquences utilisées en pilotage, ce bruit peut être très gênant dans les boucles d'asservissement des gouvernes.

Contrairement au second procédé où le filtrage était fait par le logiciel, le troisième procédé revient à l'acquisition synchrone en effectuant le retard et l'addition par des circuits discrets indépendants sur chaque voie. Le bruit en sortie est alors équivalent à celui du premier procédé.

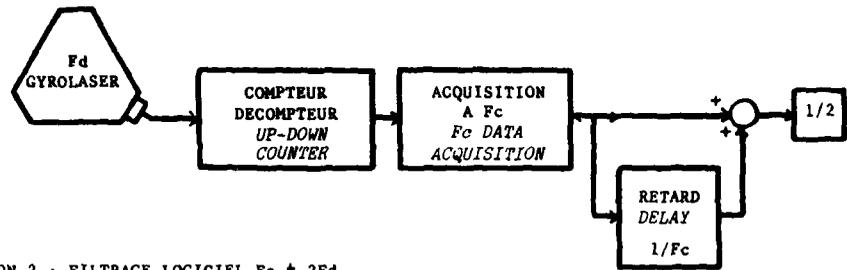
Dans le cas des procédés 2 et 3 qui font intervenir une addition entre deux acquisitions angulaires successives de durée T_c (période d'échantillonnage), le retard de datation est égal à une période d'échantillonnage T_c voisin de la demi-période d'activation $T_d/2$.

Dans le cas du procédé 1, la soustraction était faite en permanence et sans retard ; le retard de datation n'est plus que la demi-période d'échantillonnage $T_c/2$.



Solution 1

FIGURE 11
FILTRAGE DE L'ACTIVATION
OTHER MOVEMENT FILTERING



SOLUTION 2 : FILTRAGE LOGICIEL $F_c \neq 2F_d$
 SOFTWARE FILTERING $F_c \neq 2F_d$

SOLUTION 3 : FILTRAGE PAR MATERIEL $F_c = 2F_d$
 HARDWARE FILTERING $F_c = 2F_d$

FIGURE 12
 FILTRAGE DE L'ACTIVATION
 DITHER MOVEMENT FILTERING

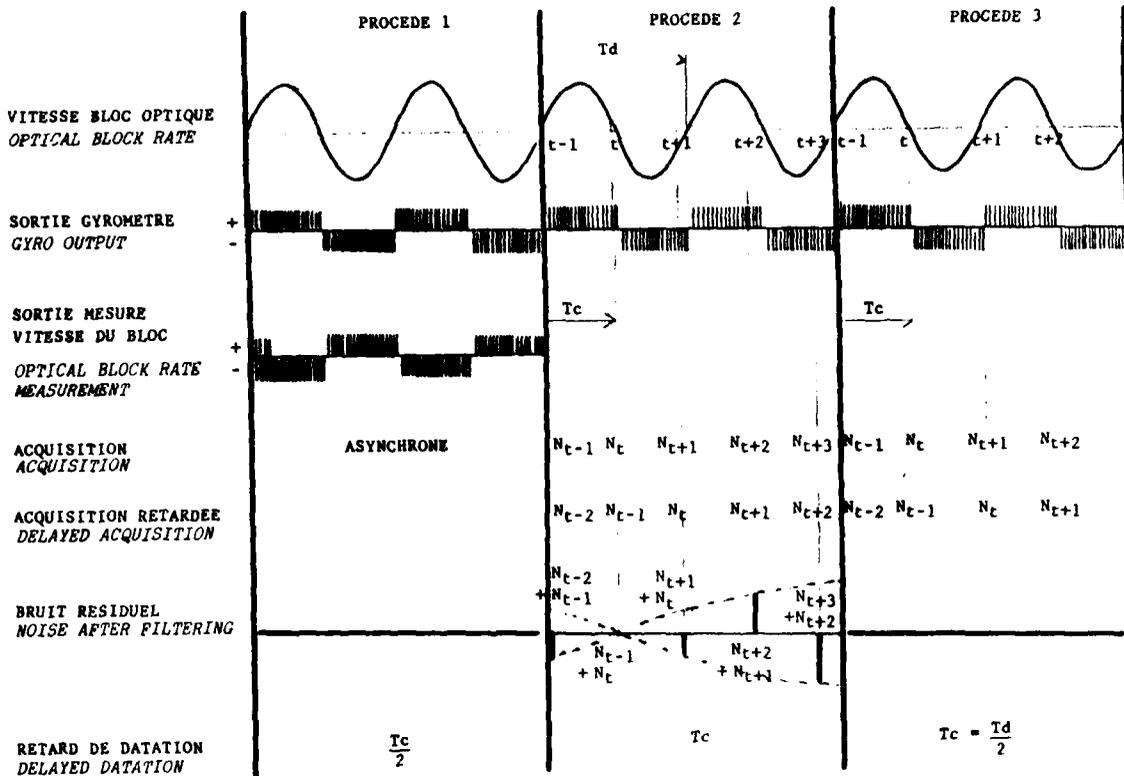


FIGURE 13
 FILTRAGE DE L'ACTIVATION A $\Omega = 0$
 DITHER FILTERING WHEN $\Omega = 0$

Si l'on veut réduire les bruits pour le pilotage, le choix reste entre les procédés 1 et 3.

A l'avantage du procédé 1 son faible retard de datation qui ne dépend que de la fréquence de calcul.

A l'avantage du procédé 3, l'absence de complexité au niveau du capteur et des circuits d'acquisition et de conversion tension fréquence.

Les circuits de retard et de comparaison supplémentaire peuvent être réalisés en technologie prédiffusée et seront donc peu onéreux.

Cette solution 3 sera retenue chaque fois qu'un retard de datation égal à la demi-période d'activation sera acceptable.

4.2. Bruits dans la chaîne gyrométrique

La chaîne gyrométrique fournissant les signaux de pilotage compte plusieurs sources de bruit. Elle est schématisée sur la figure 14.

La première source de bruit est, bien entendu, la marche au hasard du gyrolaser dont la densité spectrale qbb s'exprime en $^{\circ 2}/h$.

La seconde source de bruit correspond au bruit aléatoire imposé à l'amplitude d'activation. Elle correspond à une incertitude sur la position du gyrolaser et est caractérisée par un écart type correspondant environ à $0,3 q$ où q est le poids de l'incrément du gyrolaser.

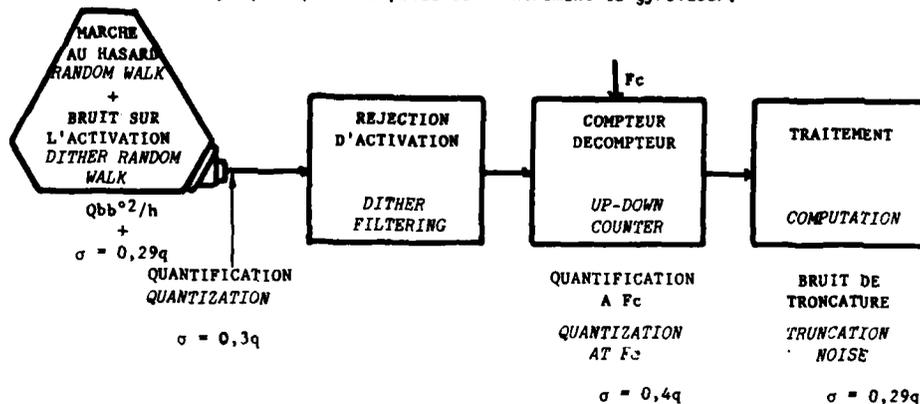


FIGURE 14
PRINCIPALES SOURCES DE BRUIT
MAIN NOISE SOURCES

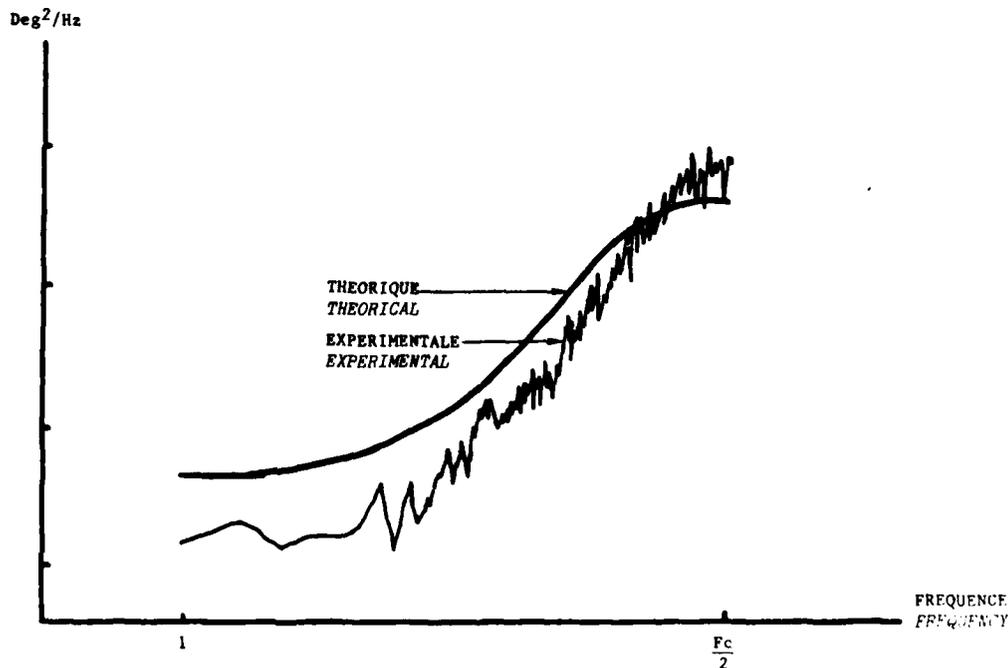


FIGURE 15
DENSITE SPECTRALE DE BRUIT
SPECTRUM DENSITY NOISE

Du fait de son type de réalisation, la réjection d'activation n'apporte pas de bruit supplémentaire. Par contre l'acquisition des incréments sur le compteur-décompteur à une fréquence F_c introduit un bruit d'échantillonnage dont l'écart type est égal à $0,4 q$.

Le traitement des échantillons peut également introduire un bruit de troncature dont l'écart type sera $\sigma = 0,29 q$.

Les calculs permettant d'établir la densité spectrale du bruit résultant de ces cinq sources de bruit sont trop longs pour être reproduits dans cet exposé.

Ils montrent que la densité spectrale de bruit croît en fonction de la fréquence jusqu'à un maximum voisin de la moitié de la fréquence d'échantillonnage F_c .

Ils confirment également que l'influence de la marche au hasard est prépondérante aux basses fréquences comme le laissait supposer la figure 5.

Des essais expérimentaux ont été faits pour vérifier l'analyse théorique. Les deux courbes de la figure 15 permettent de comparer les courbes théoriques et pratiques.

L'écart sur les basses fréquences provient de ce que le gyrolaser utilisé avait une marche au hasard beaucoup plus faible que celle prise pour faire le calcul théorique.

L'ensemble des mesures de bruit réalisé sur les systèmes inertiels pour missiles ont montré que les bruits créés aussi bien par le gyrolaser de 12 cm retenu pour ces équipements que par les traitements utilisés étaient compatibles d'un emploi sur les missiles les plus performants.

5. PERSPECTIVES

Si des blocs inertiels utilisant des gyrolasers de 12 cm sont disponibles pour équiper les missiles tactiques moyennes portées dans les dix prochaines années, on peut examiner dès maintenant dans quelle direction vont devoir évoluer ces matériels.

Il est bien évident que la demande des constructeurs de missiles ira toujours vers une diminution des encombrements et une augmentation des domaines de mesure pour réaliser des missiles plus petits et plus maniables.

Les spécifications de dérive varieront vraisemblablement peu et l'effort sera donc à porter pour les gyrolasers sur la taille sans pour autant perdre sur le niveau de bruit.

Dépendant de la zone aveugle et de la quantification, le bruit est lié à la taille des gyrolasers.

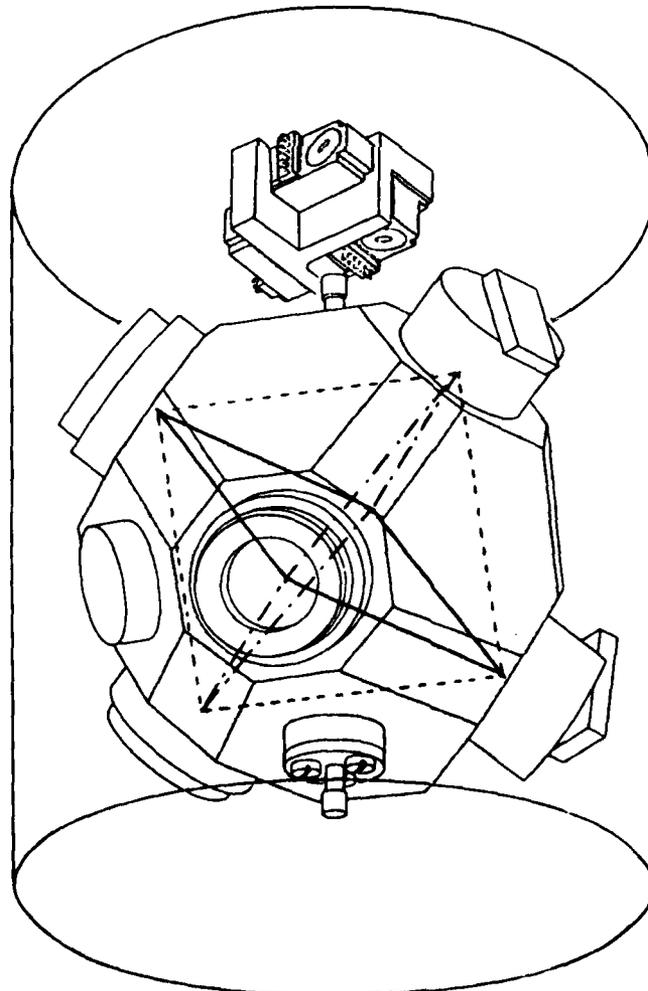


FIGURE 16
GYROLASER TRIAXE
TRIAxis LASER GYRO

Dans l'avenir, pour réduire cette taille, des améliorations sont possibles :

- d'une part, l'incrément d'angle en sortie du gyrolaser peut être réduit en utilisant des circuits électroniques adaptés pour traiter les franges d'interférence. Dans ce cas, même avec des gyrolasers de 6 cm de périmètre, le bruit de quantification peut être amélioré.
- d'autre part, les progrès incessants faits dans la technologie des miroirs de gyrolaser permettront encore de diminuer la zone aveugle et donc la marche au hasard. La taille des gyrolasers pourra donc être diminuée en même temps que la taille des électroniques associée diminuera avec les progrès de l'électronique.

Une autre voie d'évolution se situe pour les gyrolasers dans la réalisation de capteurs triaxiaux pour lesquels des travaux de développement sont en cours.

De tels capteurs plus complexes certes que les capteurs monoaxe permettront de réduire l'encombrement global tout en conservant un périmètre supérieur : 14 à 16 cm par exemple pour des parcours carré imbriqués (figure 16).

Ces capteurs ouvrent trois perspectives :

- Une réduction d'encombrement
- Une réduction des bruits pour le pilotage
- Enfin, une réduction des coûts puisqu'un capteur triaxe à 3 parcours carré n'utilise que 6 miroirs et un seul mécanisme d'activation.

Dans l'une ou l'autre de ces voies, il sera possible de développer dans les années qui viennent des ensembles inertiels à gyrolaser dont le volume n'excédera pas 1,5 l, qui seront utilisables pour une très grande gamme de missiles et dont le prix sera rendu très compétitif par l'effort d'industrialisation fait sur les gyrolasers et sur la microélectronique.

UTILISATION D'UN MAGNETOMETRE AUTOCOMPENSE DANS UN SYSTEME DE NAVIGATION ECONOMIQUE POUR HELICOPTERE

by

J.L. Roch, J.C. Goudon and Ph. Chaix
Société Crouzet
rue Jules Vedrines 25
BP 1014
F-26027 Valence, France

1. INTRODUCTION

L'analyse des missions d'un hélicoptère armé a montré le besoin d'installer sur celui-ci un système de navigation. En effet, pour assurer sa sécurité et donc ne pas être détecté par les systèmes ennemis, l'hélicoptère armé doit voler en suivi de terrain à une hauteur inférieure ou égale à 50 m et à une vitesse variant dans tout le domaine d'utilisation. Il est donc difficile de se repérer pendant ces phases de vol dites "tactiques" d'où la nécessité d'avoir un système de navigation sur l'hélicoptère (précision du système souhaité).

1.1. Présentation et analyse de différents systèmes de navigation autonome

Il existe plusieurs sortes de systèmes de navigation autonome (sans aide extérieure) :

- Centrale inertielle
- Système de navigation hybride (inertie + référence de vitesse sol)
- Système composé :
 - . d'un capteur de cap magnétique
 - . d'une centrale de verticale
 - . d'un capteur de vitesse sol
 - . d'un calculateur

Ce système de navigation permet d'utiliser deux types de cap d'origine magnétique :

1.1.1. Cap gyromagnétique

Cet appareil est composé de deux sous-ensembles : une "flux-valve" et un compas gyromagnétique. Le compas gyromagnétique (gyroscope) est asservi à long terme sur la référence magnétique de la "flux-valve" afin de compenser les différentes dérives du gyroscope (dérive propre, rotation terre, convergence des méridiens).

La "flux-valve" est un appareil pendulé qui permet de mesurer le champ magnétique horizontal (relatif à la verticale apparente) et qui doit donc être compensé magnétiquement à cause des perturbations magnétiques dues à l'hélicoptère.

Cette compensation est délicate car son automatisation est pratiquement exclue (nécessité d'effectuer la compensation en vol en statique) et elle nécessite un outillage coûteux (chercheur de Nord, théodolite,...).

De plus, lorsque le vol est agité (vol tactique) de nombreuses erreurs de cap apparaissent qui sont dues :

- à la vanne de flux (erreurs dues à la pendularité)
- à la surveillance magnétique (saturation de l'asservissement, coupures de surveillance fréquentes dues aux attitudes et aux accélérations)
- au gyroscope directionnel (fonctionnement fréquent en directionnel, erreurs de cardan importantes, sauf pour les centrales bi-gyroscopiques).

1.1.2. Cap d'origine magnétométrique

Le magnétomètre est un capteur qui permet de mesurer les trois composantes du champ magnétique ambiant suivant un repère orthonormé.

Les composantes ainsi mesurées sont ensuite compensées des différentes perturbations magnétiques, puis projetées en repère long-traverse (repère horizontal) d'où on en déduit le cap magnétique.

Le magnétomètre présente les avantages suivants :

- il ne nécessite pas l'emploi d'un gyroscope directionnel car le magnétomètre n'est pas pendulé, donc pas soumis aux mêmes erreurs que la "flux-valve" pendant les phases accélérées.
- il est d'une utilisation simple et pratique grâce notamment au principe de la compensation automatique en vol qui supprime les contraintes liées à une calibration délicate sur la machine et qui permet une surveillance du bon fonctionnement du capteur et de la stabilité des perturbations magnétiques.
- le coût du système est plus modeste car on n'utilise pas de compas gyromagnétique (gyroscope).

1.2. Choix du capteur magnétique

C'est le magnétomètre statique trois axes que nous avons retenu comme moyen de mesure de la référence magnétique car il permet :

- une compensation plus facile et automatique
- un coût plus faible
- une fiabilité plus importante car il ne comporte pas d'élément mobile.

2. PRESENTATION DU SYSTEME DE NAVIGATION

Le système est composé de :

- un calculateur NADIR (CROUZET) qui assure :
 - . l'acquisition des capteurs
 - . les différents traitements numériques (compensation, calcul du cap, entretien de la navigation,...)
 - . la visualisation et l'entrée des données (PCV)
- un cinémomètre Doppler type RDN 80 B (ESD) qui mesure dans un repère lié à l'avion les composantes de la vitesse sol
- un gyroscope de verticale GV 76-1 (SFIM) qui fournit les informations d'attitude (roulis et assiette longitudinale)
- un magnétomètre statique 3 axes (CROUZET)

Ce capteur est constitué de trois sondes dont les axes définissent un trièdre trirectangle de mesure, harmonisé avec les axes avion. Chaque sonde est constituée d'un noyau en matériau magnétique entouré d'un bobinage co-axial parcouru par un courant alternatif haute fréquence. Ce bobinage présente, lorsque le noyau est soumis à un champ magnétique extérieur, des dissymétries de tension à ses bornes : ces dissymétries sont détectées et annulées par un courant continu qui est donc proportionnel à la composante du champ mesuré par la sonde. Le magnétomètre fournit donc trois tensions continues proportionnelles aux projections du champ magnétique local dans un repère lié au porteur.

- Indicateur de navigation type 152 (CROUZET)

Cet instrument de bord permet la visualisation des informations suivantes :

- . cap magnétique sur une rose
- . route magnétique sur un index
- . gisement d'une radio-balise
- . distance au but de destination sur un compteur
- . direction du but de destination sur une aiguille
- . alarmes sur trois drapeaux (cap, gisement, distance).

L'entretien de la position présente est assuré par intégration des composantes Vitesse Nord et Vitesse Est qui sont obtenues par projection des vitesses doppler à l'aide des informations d'attitude (roulis et assiette longitudinale fournis par le gyroscope de verticale) et de l'information de cap géographique (obtenu par cap magnétique + déclinaison).

3. ELABORATION DU CAP MAGNETIQUE

3.1. Calcul du cap magnétique

A partir des trois composantes du champ magnétique terrestre (H_x , H_y , H_z) et des informations (θ , ψ) on peut calculer les composantes long et travers du champ d'où on en déduit le cap magnétique (voir figure 2 pour la définition des angles et la figure 3 pour le principe de calcul)

$$H_L = H_x \cos\theta + H_y \sin\theta \sin\psi + H_z \sin\theta \cos\psi$$

$$H_T = H_y \cos\psi - H_z \sin\psi$$

$$\text{d'où } C_m = \text{Arctg} \left(- \frac{H_T}{H_L} \right) + k\pi$$

L'ambiguïté de 180° sur le cap due à la fonction arctangente est facilement levée par l'examen du signe de H_L .

3.2. Problème particulier de la compensation

Le champ magnétique mesuré par le magnétomètre 3A n'est pas exactement le champ magnétique terrestre car celui-ci est perturbé par l'environnement.

3.2.1. Rappels sur les perturbations

- perturbations de fers durs : elles traduisent les effets des aimantations rémanentes dans les matériaux ferromagnétiques et les champs créés par les courants continus. Ces perturbations peuvent se modéliser par un champ additionnel lié aux axes du porteur :
- perturbations de fers doux : elles traduisent les déformations des lignes de champ magnétique dues à l'inhomogénéité de perméabilité magnétique du porteur. Elles se modélisent par un champ perturbateur qui est le produit d'un tenseur [K] par le champ magnétique terrestre.

En conclusion, les mesures du champ magnétique sur le porteur peuvent se modéliser sous la forme suivante :

$$\begin{pmatrix} H_{mx} \\ H_{my} \\ H_{mz} \end{pmatrix} = \begin{pmatrix} k_{xx} & k_{xy} & k_{xz} \\ k_{yx} & k_{yy} & k_{yz} \\ k_{zx} & k_{zy} & k_{zz} \end{pmatrix} \cdot \begin{pmatrix} H_x \\ H_y \\ H_z \end{pmatrix} + \begin{pmatrix} b_x \\ b_y \\ b_z \end{pmatrix}$$

Champ mesuré
tenseur de perturbation de fers doux
champ terrestre
vecteur de perturbation de fers durs

3.2.2. Principe de la compensation

S'il n'y avait pas de perturbations magnétiques, on s'apercevrait que si l'hélicoptère pouvait prendre toutes les orientations possibles, l'extrémité du champ magnétique se déplacerait sur une sphère dans les axes avion (à condition de rester dans un même lieu car le module du champ varie avec la position).

Du fait des perturbations magnétiques (fers durs et fers doux), l'extrémité du champ magnétique mesuré décrit un ellipsoïde.

Le but de la compensation est donc de calculer les caractéristiques de cet ellipsoïde . son centre . sa forme.

Or, on ne peut pas décrire l'ellipsoïde complètement (pas de vol sur le dos) par conséquent on a choisi un certain nombre de figures caractéristiques qui apportent suffisamment d'informations pour pouvoir calculer son équation.

Les figures choisies sont plusieurs virages (de 360° en cap) à roulis constant ($\psi = +45^\circ$, $\psi = -45^\circ$, $\psi = +30^\circ$, $\psi = -30^\circ$).

Afin de mieux voir la partie de l'ellipsoïde qui est parcourue, on a représenté sur la figure 4 la perspective isométrique de la figure décrite par l'extrémité du vecteur champ lorsque l'hélicoptère fait des 360° à $\psi = \pm 45^\circ$, $\psi = \pm 30^\circ$, $\psi = \pm 10^\circ$.

Une fois l'ellipsoïde déterminé, on détermine alors les paramètres de fers durs et fers doux (\vec{b} et [K]) à partir de son équation qui permettent de transformer l'ellipsoïde en sphère. L'opération de compensation consiste donc à associer chaque point de l'ellipsoïde à un point de la sphère.

Le déroulement d'une autocompensation est alors le suivant :

- Le pilote engage le mode autocompensation.
- Il effectue ensuite la série de figures préconisées jusqu'à ce que le calculateur NADIR lui indique de s'arrêter (convergence de l'algorithme après 4 à 5 minutes de vol).
- A l'issue de cette autocompensation, les paramètres \vec{b} et $[K]$ sont figés et stockés en mémoire permanente afin d'être utilisés pour tous les vols suivants.

De plus, afin d'assurer une surveillance du capteur ou de déterminer si l'état magnétique de l'hélicoptère a changé (changement de pièces, embarquement de charges magnétiques) on continue pendant les vols suivants à réactualiser l'équation de l'ellipsoïde et lorsqu'on constate qu'il y a une trop grande différence entre l'ellipsoïde actuel et l'ellipsoïde figé, on déclare qu'il y a un problème. (Le pilote décide alors s'il doit réeffectuer ou non une nouvelle autocompensation).

3.3. Calcul du cap magnétométrique

Un diagramme fonctionnel de la chaîne de cap est donné figure 5.

On effectue un regroupement des deux fonctions présentées précédemment (calcul du cap et compensation).

4. EXPERIMENTATION DU SYSTEME DE NAVIGATION AUTONOME AVEC MAGNETOMETRE

4.1. Introduction

Le système de navigation autonome a été installé sur deux types d'hélicoptère français :

- l'hélicoptère PUMA SA 330
- l'hélicoptère GAZELLE SA 342.

et essayé par le Centre d'Essai en Vol français et par l'Armée de Terre française.

4.2. Essai du système de navigation sur PUMA

4.2.1. Présentation de l'installation

L'hélicoptère PUMA est un hélicoptère de la classe 6 à 7 tonnes avec une vitesse de croisière de l'ordre de 270 km/h.

Le magnétomètre a été installé dans la queue de l'hélicoptère dans un endroit magnétiquement propre et pas trop perturbé.

4.2.2. Résultats obtenus

Cette première évaluation qui a eu lieu en 1979 a permis de montrer la validité du principe d'autocompensation et la capacité du cap déterminé à satisfaire une performance en navigation correcte.

La représentation sur la figure 8 des erreurs de cap avant et après compensation montre l'efficacité de l'autocompensation qui permet d'obtenir des erreurs de cap inférieures à 0,5°.

On n'avait pas pu tirer des conclusions sur les performances de ce système en navigation car trop peu de mesures avaient été effectuées.

4.3. Essai du système de navigation sur GAZELLE

4.3.1. Présentation de l'installation

L'hélicoptère GAZELLE est un hélicoptère de la classe 2 tonnes avec une vitesse de croiseur de l'ordre de 250 km/h.

Le magnétomètre a été installé en arrière du doppler sous l'emplanture de queue de l'hélicoptère.

4.3.2. Résultats obtenus

Les essais du système ont été menés en 1982 et 1983, en deux étapes successives :

- . une phase de mise au point, suivie d'une courte période d'évaluation a été menée au C.F.V. de Brétigny sur Orge.
- . une phase plus longue d'évaluation opérationnelle.

Après une période de mise au point, les essais ont montré un bon comportement de l'autocompensation, réalisée en moyenne en 4 minutes, à l'issue d'une procédure comportant en tout quatre à cinq virages à inclinaison différente ($\pm 30^\circ$, $\pm 45^\circ$, $\pm 60^\circ$).

A titre d'exemple, la courbe de régulation présentée figure 7 montre l'efficacité de l'autocompensation : erreur de cap comprise entre $-0,5$ et $+0,4$ degrés.

L'évaluation opérationnelle du système a été effectuée d'une part pour l'autocompensation, d'autre part pour la navigation.

Autocompensation

Une vingtaine d'autocompensations ont été réalisées en tout, avec des équipages différents les uns des autres et non spécialisés.

Ces éléments sont particulièrement importants puisqu'ils correspondent bien aux conditions opérationnelles.

L'autocompensation est réalisée par une procédure comportant 4 ou 5 virages, nécessitant 4 à 5 minutes de vol.

Les réactions enregistrées démontrent une entière satisfaction des utilisateurs à l'égard de la méthode, ainsi que des résultats qui se sont révélés très stables.

Résultats des vols de navigation (transit)

Deux cent quatre vingt huit (288) branches ont été réalisées, d'une longueur comprise entre 25 et 45 km à des vitesses de l'ordre de 180 km/h à 220 km/h.

Les résultats sont consignés sur la courbe figure 8. Ils correspondent à une performance globale de : 1,7 % de la distance parcourue dans 95 % des cas.

Résultats des vols tactiques

Quatre vingt branches de vol tactique d'une durée de 15 minutes environ ont été réalisées.

Les résultats sont consignés sur la figure 12 où la performance globale est de : 450 m par quart d'heure de vol dans 95 % des cas.

5. CONCLUSION

Les résultats obtenus au cours des différents essais ont permis d'améliorer progressivement les performances de la fonction "cap autocompensé par magnétomètre statique".

Au vu des divers résultats obtenus et officiellement constatés, le magnétomètre statique constitue une source de cap magnétique autocompensé, de haute qualité, libérant l'utilisation des contraintes liées à la compensation périodique de la machine et apte à réaliser une navigation Doppler de haute précision, conforme aux exigences opérationnelles de l'Armée française soit par exemple avec un GV 76 et un RDN 80 B.

| | |
|------------------------------|-------------------|
| 2 % de la distance parcourue | sur terre, à 95 % |
| 500 m par quart d'heure | |

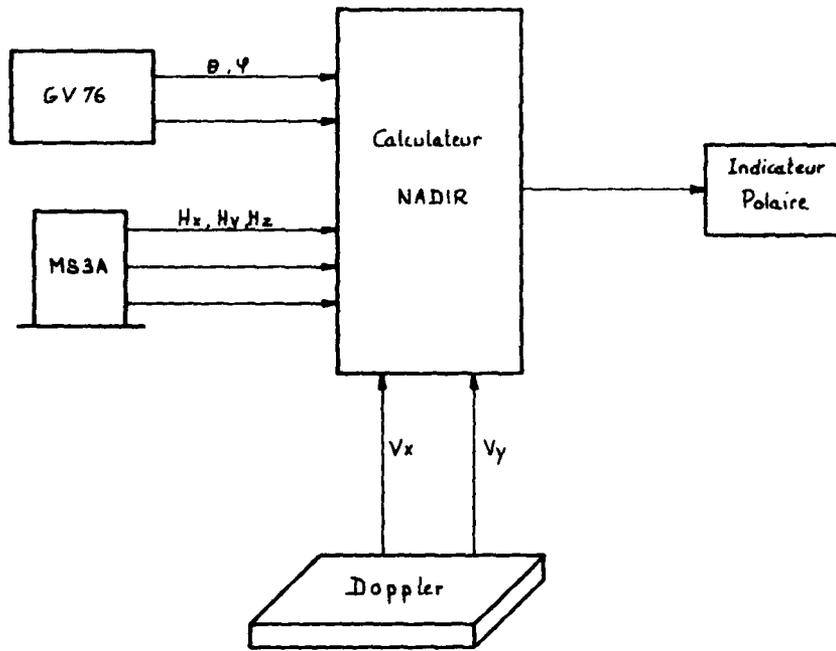


FIGURE 1 : SCHEMA DU SYSTEME

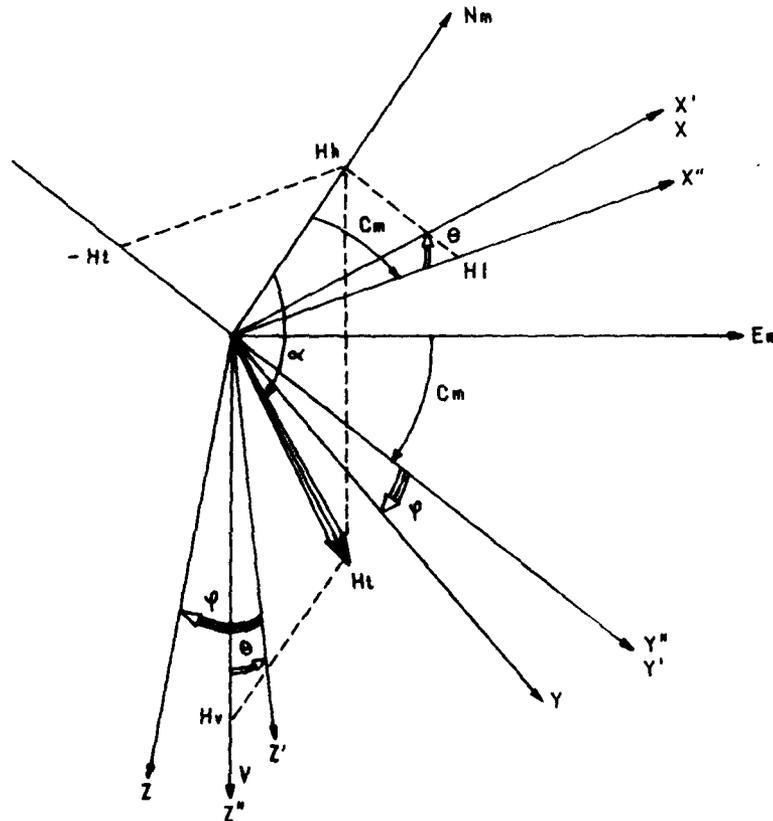
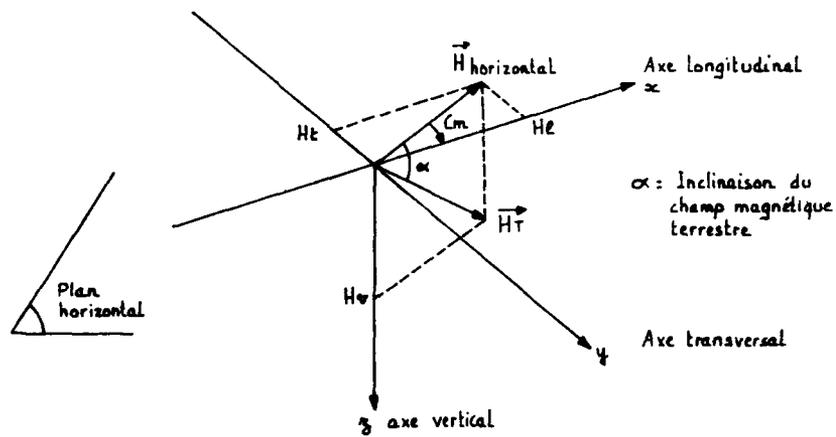


FIGURE 2 : DEFINITION DES ANGLES



α : Inclinaison du champ magnétique terrestre.

$$C_m = \text{Arctg} \left(\frac{-H_v}{H_h} \right) + \pi$$

FIGURE 3 : PRINCIPE DE CALCUL

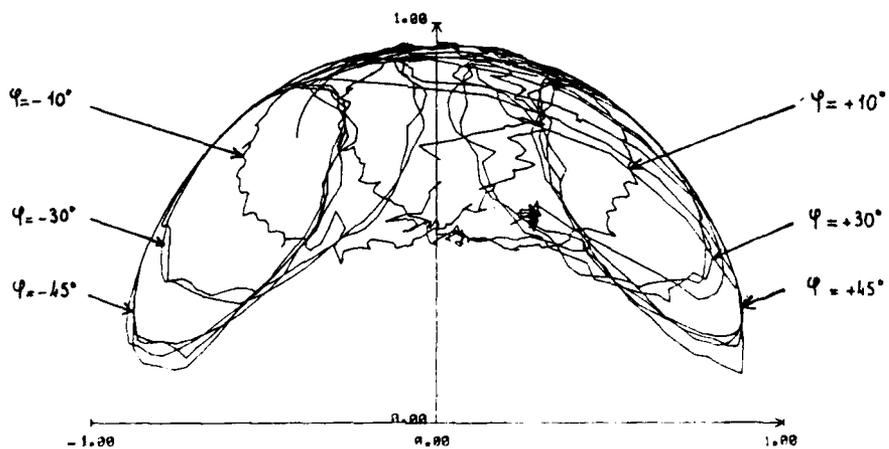


FIGURE 4 : PERSPECTIVE ISOMETRIQUE DE LA FIGURE DECRITE PAR LE VECTEUR CHAMP EN VIRAGE

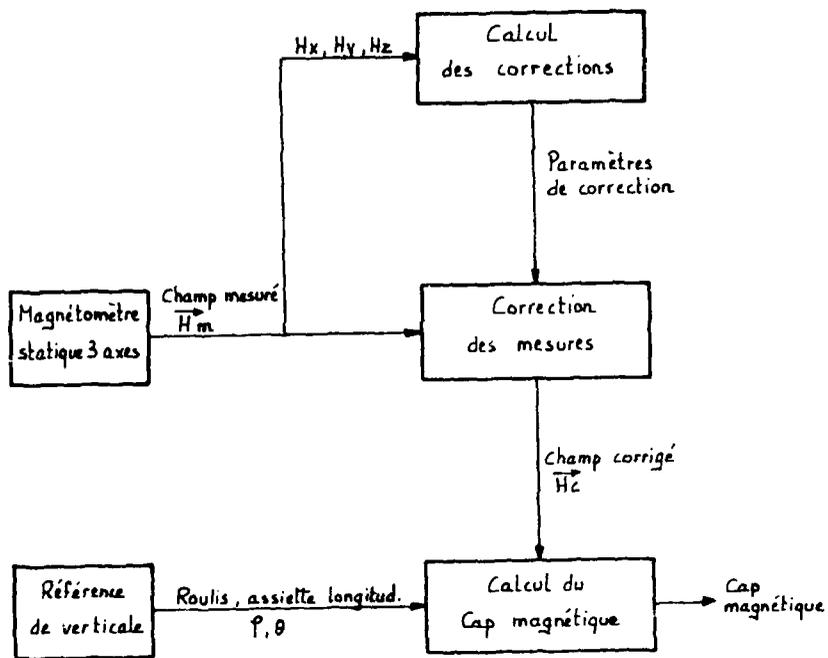


FIGURE 5 : DIAGRAMME FONCTIONNEL.

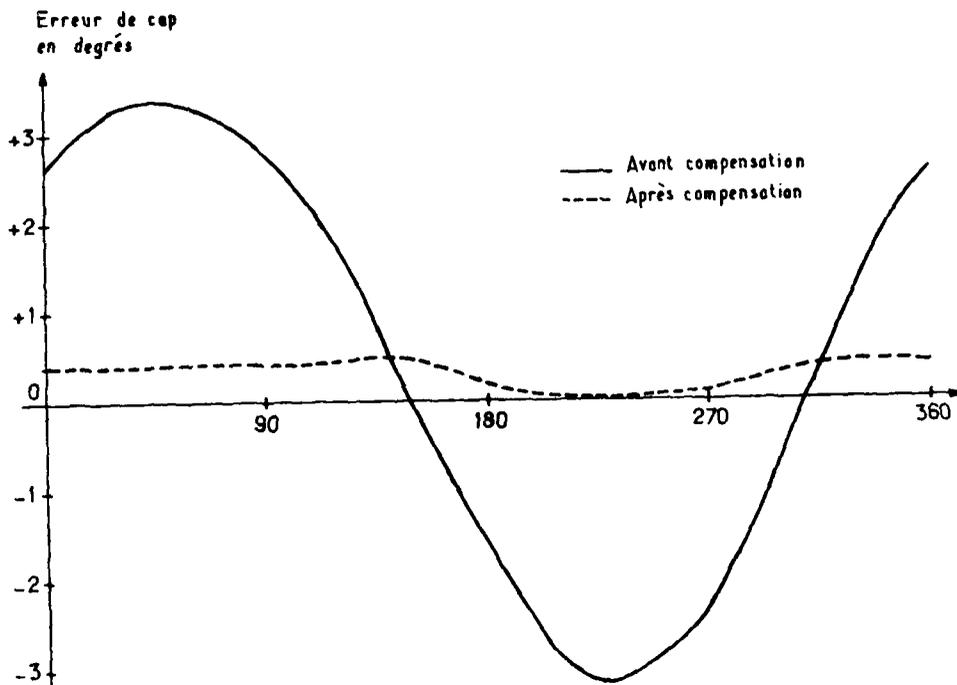


FIGURE 6 : COMPENSATION D'UMA

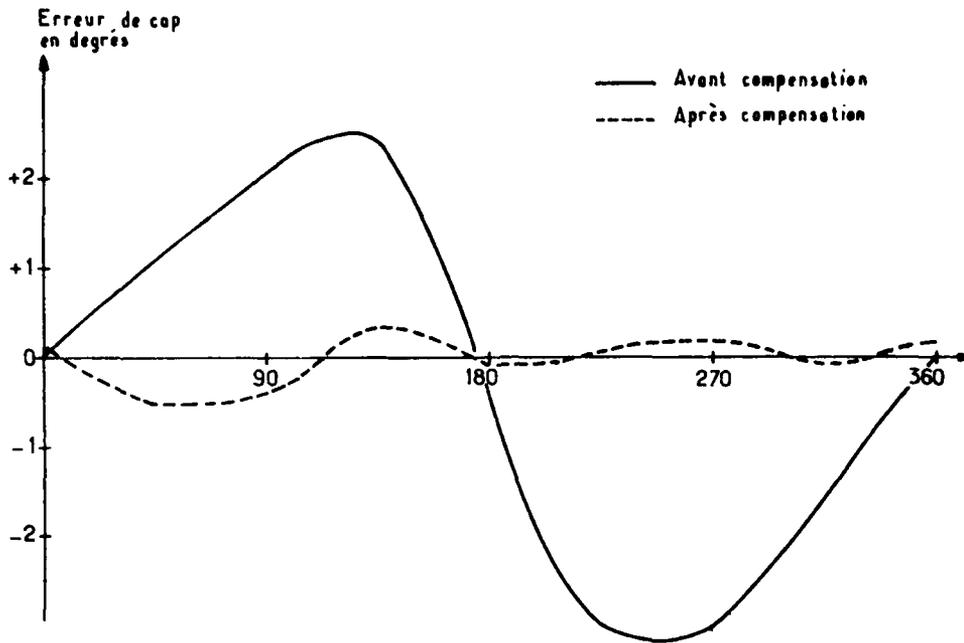


FIGURE 7 : COMPENSATION GAZELLE

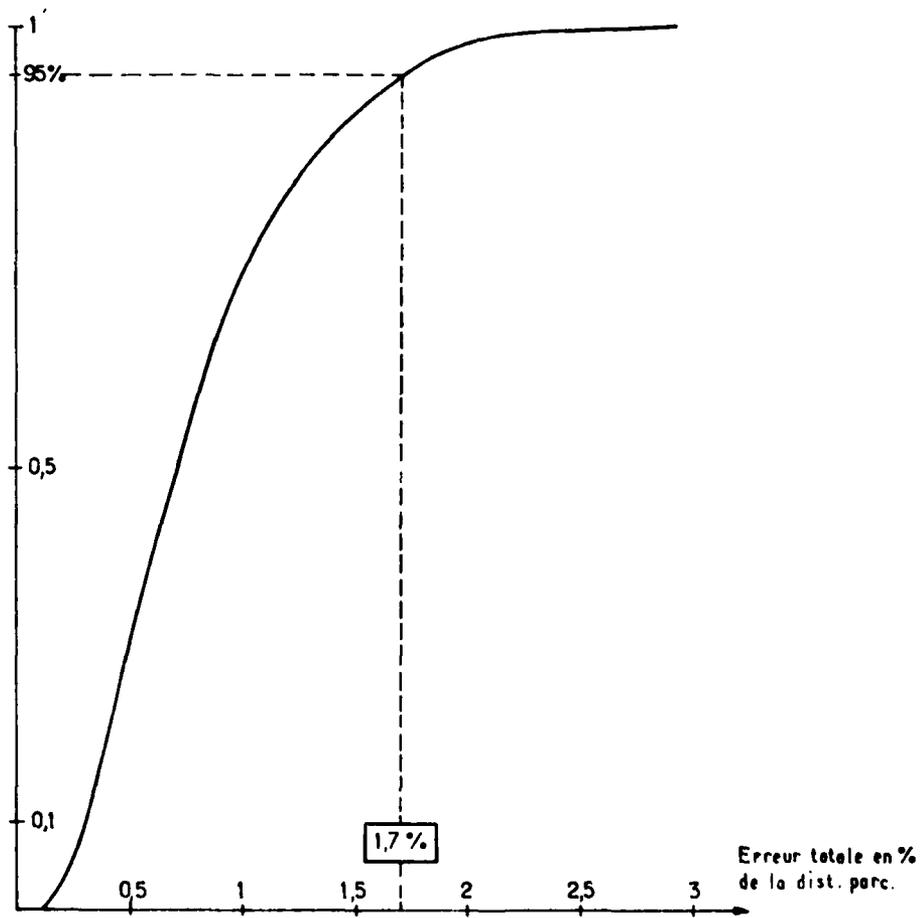


FIGURE 8 : RESULTAT NAVIGATION TRANSIT

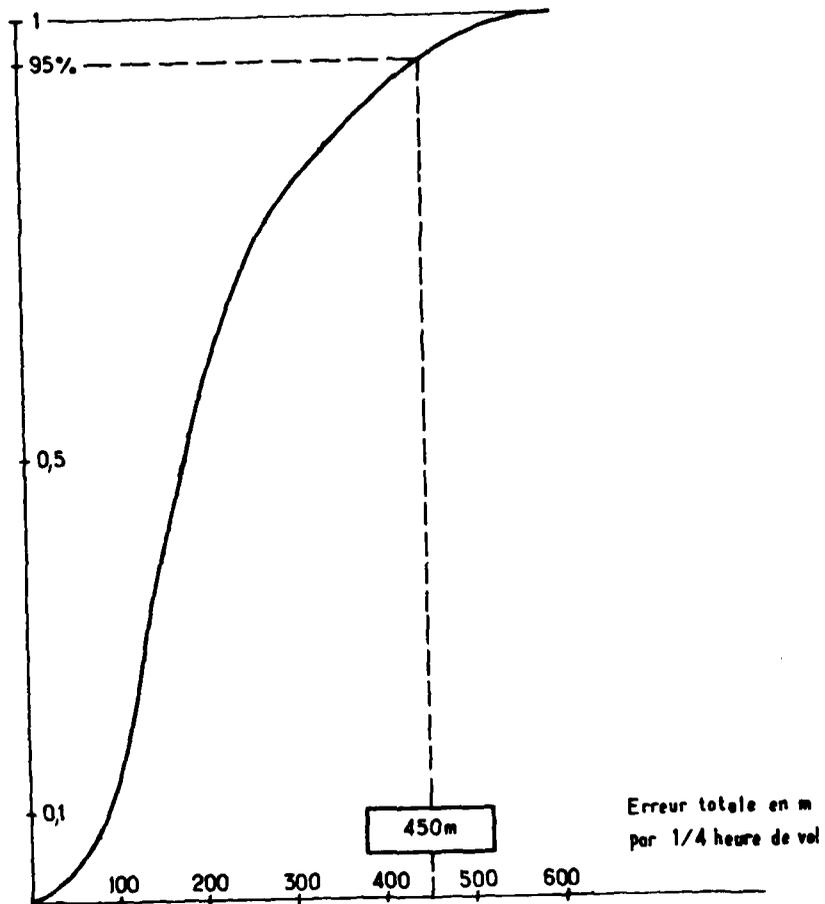


FIGURE 9 : RESULTATS EN VOL TACTIQUE

A COST-EFFICIENT CONTROL PROCEDURE
FOR
THE BENEFIT OF ALL AIRSPACE USERS

by

André Benoît* and Sip Swierstra

European Organisation for the Safety of Air Navigation
EUROCONTROL
Rue de la Loi, 72, B-1040 Bruxelles

SUMMARY

Provided adequate directives are generated by an advanced ATC system, the use of onboard computers in guidance, navigation and control systems will make it possible for flights to be conducted in such a way as to agree closely with optimisation directives irrespective of the particular criterion used and of whether the context is a civil or military one. In addition, the air/ground control coordination will be efficiently performed by means of automated digital data communications. Advances in the relevant technologies can be expected to continue and as such advances are brought into actual use the control environment will change accordingly. However, the new facilities and equipment involved will probably be expensive to purchase and run, and much attention will need to be given to the cost aspect of their utilisation.

We are therefore proposing, in this context, a control procedure for conducting time-of-arrival-constrained flights in an economic manner. This procedure is intended to be compatible with present-day voice communications (human or synthetic) although it is primarily designed to be used in conjunction with future automated digital data communications.

The procedure is applicable to the transit of flights through extended terminal areas such as are considered in connection with Zone-of-Convergence-type systems, and the final approach phase is accordingly integrated with the en route descent, cruise and (possibly) climb phases, or appropriate parts thereof.

The paper contains a detailed description of the procedure, together with brief summaries of the tests conducted in present and simulated future environments to assess its efficiency, and then sets out the results obtained to date and analyses them in terms of 4-d navigational accuracy and operational effectiveness.

1. INTRODUCTION

An air traffic control system rethought in the light of present economic constraints, in particular one enabling duly selected trajectories to be followed, certainly brings about an appreciable reduction in the overall cost of flights (Refs. 1, 2). At European medium-to-high-density traffic airports, such as Brussels and London, proper integration of en-route and approach control can be expected to lead to a reduction of some 10 to 20 per cent in the fuel actually burned by the inbound traffic in the zones of convergence including and surrounding these terminal areas (Ref. 3).

Efforts have been made in several quarters in Western Europe to determine optimum trajectories for each individual aircraft in relation to overall traffic, with emphasis on specific criteria, such as maximum use of available landing capacity, minimum cost of total flights inbound to a given terminal area, or minimum deviation from the profiles requested by operators (Refs. 4 to 10.)

In the case of all the approaches proposed or described in References 4 to 10, the trajectory of each inbound flight is normally time-of-arrival-constrained as a result of the optimisation process applied to the overall traffic. Accordingly, for a Zone-of-Convergence-type control system, it is essential to select an appropriate trajectory for each individual aircraft and control it with a high degree of accuracy. With flight management computer systems configured for 4-d navigation, the conduct of the flight could be controlled within a few (2-5) seconds (Ref. 11). But what could be achieved with present modes of operation, and currently available navigation equipment and control facilities?

Developments have been undertaken in order to define and assess a ground/air coordinated control procedure which would permit accurate control under present-day operational conditions (Ref. 12). Obviously, for the efficient use of such a procedure the ground-based control would require adequate automated aids, underpinned by a detailed knowledge of the aircraft operation and performance characteristics. The successive control actions resulting from landing time slot allocation, and prediction reliability, necessitate an air/ground control procedure program operated on-line, for raising, confirming and updating the requisite control directives, including ground/air control message generation.

The present paper describes the ground/air coordinated procedure proposed for the accurate control of trajectories and gives a summary of the tests conducted to date. A subsequent dissertation will present the structure and mode of operation of a suitable program to raise the successive control directives.

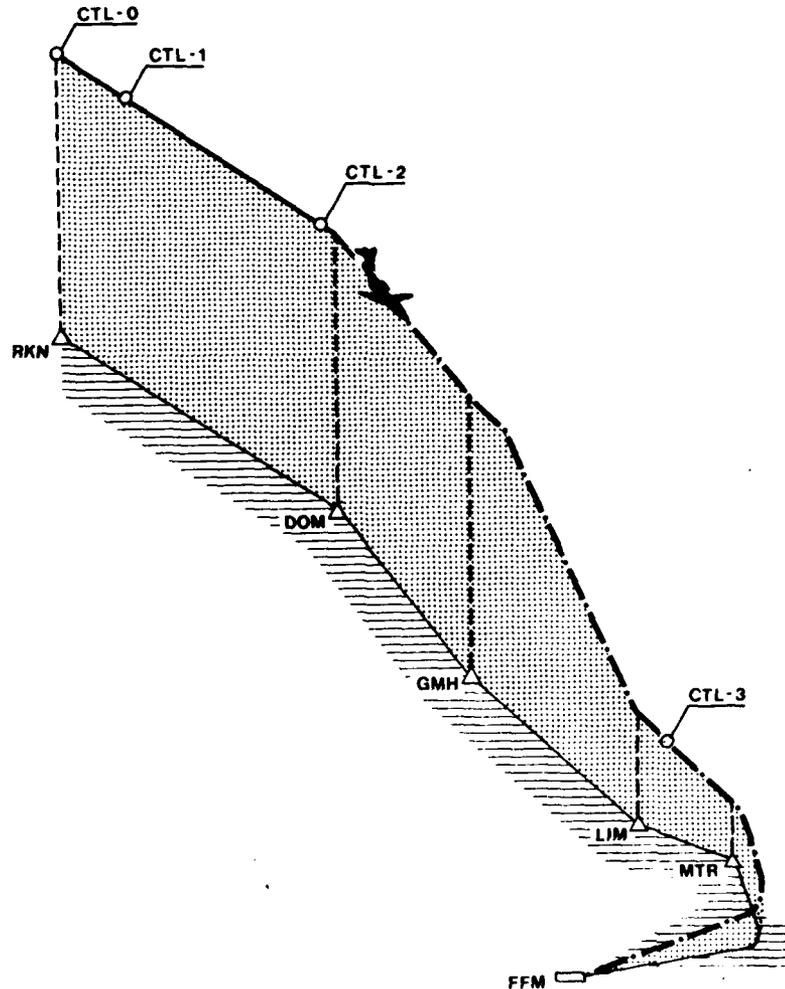
*Dr. A. Benoît is also *Chargé de cours*,
Faculté des Sciences Appliquées,
Université Catholique de Louvain,
B-1348 Louvain La-Neuve.

2. GROUND/AIR CONTROL PROCEDURE

The definition of such a procedure was intended to establish the directives needed to control accurately the trajectory of an aircraft down to the runway, the time of arrival being constrained. The essential requirements include compatibility with on-line operation, in particular with present navigation capability and voice (human or synthetic) communications, easy adaptation to automated digital ground/air communications, reliability in terms of messages generated and suitability to cope with the expected range of aircraft types and a variety of on-board control, guidance and navigation facilities.

2.1. Individual aircraft-trajectories versus air traffic

In a ZOC-type system, the ZOC sequencer/scheduler proposes an initial landing slot time when the aircraft enters the system. When doing so, it may update the time of arrival of other aircraft, possibly modifying the landing sequence determined previously. This must obviously be consistent with (a) individual flight requirements and (b) the optimization criterion applied to the overall traffic, the instantaneous and local constraints and the control variables which are or remain available at that moment. To keep the number of ground/air control actions within reasonable limits, the aim is, wherever possible, to associate only one control directive to each basic phase of the inbound flight, namely cruise or part thereof, en-route descent, and final descent and landing. The relevant actions and messages will be described in subsequent paragraphs.



SCHEMATIC CONFIGURATION OF AN INBOUND FLIGHT TO FRANKFURT

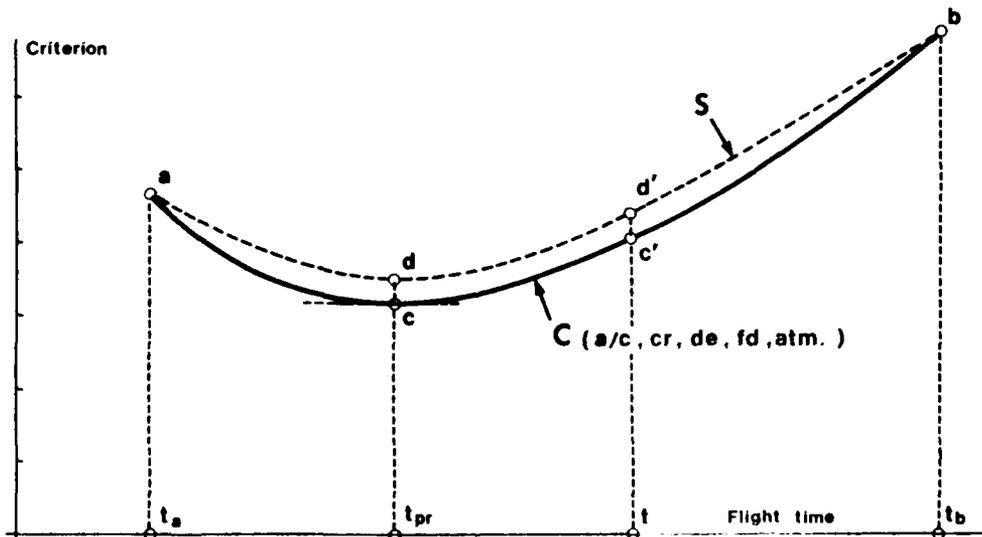
Figure 2-1

The ground sequencer/scheduler derives for each individual aircraft the trajectory which meets an overall criterion. In general, the individual operator's criterion will be the same as the criterion used to control the overall traffic, in particular where "cost of flight" is used. Nevertheless, individual operators' criteria may be altered, even on-line, and the ground optimization process will make provision for this. One possible way of coping readily with on-line requests would consist in introducing them in terms of preferential profiles and minimizing the overall deviation from such profiles. Whatever the criteria implemented by the ground-based system, the automated sequencer/scheduler should propose an optimum or sub-optimum landing slot and an associated trajectory to meet this slot. This paper concerns itself with the "implementation" and accurate control of such a trajectory, regardless of the criterion leading to its definition. However, it should be understood that the determination of such a trajectory results from the integration of management and control over the complete flight conducted within the Zone-of-Convergence geographical area, including cruise, en route descent and final descent until touch down.

2.2. Definition of an optimum time-of-arrival-constrained trajectory

The trajectory proposed for the flight from entry to touch-down, say from Rekken to Frankfurt airport as in the schematic configuration shown in Figure 2-1, will normally be composed of three main phases, namely the cruise phase, the en route descent and the final descent to touch-down. The en-route component will essentially be controlled by cruise and descent speeds expressed in terms of Mach/CAS profiles, while the control of the final descent will also depend on the local geographical configurations and may, in addition, include the precise definition of the final path as a control parameter. Clearly, any transition between successive phases (following the entry of the aircraft to the ZOC cruise conditions, from cruise to en route descent, from en route to final descent) will need to be defined accurately, with no ambiguity as to the resulting trajectory (in present ATC operation, for instance, clearance to descent implies no particular transition to descent nor any specific descent trajectory.)

Normally, the selection of the flight characteristics should result from the criterion versus time-of-transit relationship. For the contribution of a particular aircraft, the corresponding relation (C) appears as shown schematically in Figure 2-2. The influencing parameters include the aircraft mass, the cruise altitude and speed, the en route descent speed, the descent being currently conducted at or near idling conditions, the geometrical path and speed profile for the final descent and the atmospheric, wind and temperature conditions. For given entry conditions (mass, altitude, speed) and a given route to touch-down, each point of curve C corresponds to a specific trajectory for which the overall speed profile is determined. The preferential trajectory minimizes the criterion (flight duration and criterion value are noted, t_{pr} and c , respectively, on the diagram). The ZOC-recommended trajectory requires a transit duration t and leads to the criterion value c' . With respect to the preferential speed, adjustments provide a control range which extends from c to a for advanced arrival (t_{pr} , t_a) and from c to b for delayed arrival (t_{pr} , t_b). For the en route phase from entry to, say, 5,000 ft, it is appropriate to introduce a smooth cruise-to-descent transition for which both cruise and descent speed components are expressed by the same speed indication. Usually, the relevant approximation of the criterion, curve S, is quite satisfactory, suggesting that the initial trajectory could be selected accordingly.



DEFINITION OF AN OPTIMUM TIME-OF-ARRIVAL-CONSTRAINED TRAJECTORY

Figure 2-2

2.3. Control points and A/G messages

For convenience, the flight within the integrated control area will be considered as being made up of three phases, as follows.

A preparatory phase, starting when the aircraft is notified to the ZOC Management Centre. It consists in making a prediction of the runway load based on all the information available, estimates of all entries included, and accordingly assigning, tentatively, a preliminary landing slot to the aircraft entering. It ends with the transfer of control and the ground-to-air messages acknowledging the information received prior to entry.

The en route phase, which follows. The first ground-to-air message confirms, or otherwise, cruise altitude and route, then the first ZOC message is prepared and sent to the aircraft. It is sent as early as possible after entry, and includes route, new cruise speed and the position at which to initiate transition from entry speed to new cruise speed, this position being expressed in terms of DME distance from an appropriately located station, Dortmund in the case of the configuration illustrated (Figure 2-1). A second ZOC message will be issued subsequently. This message contains the en route descent characteristics, namely, the position of the initiation of the transition from cruise to descent, noted CTL-2 in Figure 2-1, also expressed in terms of DME distance from a suitably located station (Frankfurt in the illustration), the en-route descent speed profile (Mach/CAS), confirmation of the route, possibly high-altitude holding directives and the estimated arrival conditions (time and altitude) over the next active control point (noted CTL-3 in Figure 2-1) also defined in terms of DME distance from a suitable station, Frankfurt FPM in the example of Figure 2-1.

The final descent phase, including a number of messages and active control points (CTL-3(a), (b), etc.), which will depend on the local geographical configuration and type of approach considered. This phase is critical and will be discussed in detail in a subsequent section.

The tests made to assess the feasibility of the procedure, the accuracy of the 4-d controlled trajectory, during both the en-route and final descent phases, and the range available for correction, whatever the sources of dispersion, will be summarised and discussed in subsequent sections.

3. EN-ROUTE CRUISE/DESCENT PHASE

3.1. Overview

This phase has been discussed previously (Refs. 13 to 16) and it will accordingly be only outlined in the present paper. A total of 32 flights were conducted on the route Pampus (PAM), Dortmund (DOM), Germinghausen (GMH), Limburg (LIM), Metro (MTR). The flights were conducted using aircraft flight simulators: Boeing 737 and Airbus A-300 from the Deutsche Lufthansa, Flight Simulation Department, McDonnell Douglas DC-10 from SABENA, Belgian World Airlines and Fokker FK-28 from the Dutch National Aerospace Research Laboratory. A summary of the flights conducted is given in Table 3-1. The control directives to be sent to the pilot were generated on-line. Both voice and automated digital communications were considered, and in each mode of operation the directions were sent some 30 to 60 seconds, depending on the occasion, before the required action had to be initiated.

The aims of the tests conducted covered three essential aspects, namely (a) the practical operational character of the procedure (generation, transmission and acknowledgement of the control directives), (b) the accuracy of the en-route component of the 4-d trajectory from entry into the ZOC control area to the assembly point represented by Metro, 5,000 ft, in these exercises, and (c) the range of control available prior to the final descent.

The scenario of each flight is in line with the procedure described in Section 1. The time of transit is determined at entry (at the latest some 30 to 60 seconds before CTL-1); the relevant CTL-1 position and cruise speed are accordingly sent to the aircraft. Subsequently, the en-route descent characteristics (position of the relevant control point, CTL-2, and speed profile) are also sent from 30 to 60 seconds in advance, and the aircraft proceeds to the assembly point, here Metro. There, the altitude and time are compared against the initial prediction. The procedure to be followed was explained to the pilots prior to their flights by means of a short briefing.

| FLIGHTS | ROUTE | LENGTH | AIRCRAFT | OPERATORS | PERIOD | REPORT |
|---------|---------|--------|----------|-------------------------|---------|---------|
| 5 | RKN-MTR | 140 | B737 | LUFTHANSA PILOTS | JUL. 81 | 812020 |
| 3 | RKN-MTR | 140 | A-300 | LUFTHANSA PILOTS | JUL. 81 | 812020 |
| 7 | PAM-MTR | 201 | DC-10 | SABENA CHIEF INSTRUCTOR | MAR. 82 | 822028 |
| 9 | PAM-MTR | 201 | DC-10 | SABENA CHIEF INSTRUCTOR | NOV. 82 | 832028 |
| 4 | PAM-MTR | 201 | B-737 | LUFTHANSA PILOTS | MAR. 83 | 832028 |
| 2 | PAM-MTR | 201 | A-300 | LUFTHANSA PILOTS | MAR. 83 | 832028 |
| 1 | PAM-MTR | 201 | FK-28 | NLR PILOT | DEC. 83 | 84E1/TN |
| 1 | GMH-MTR | 68 | FK-28 | NLR PILOT | DEC. 83 | 84E1/TN |

SUMMARY OF 32 EN-ROUTE CRUISE/DESCENT TESTS CONDUCTED ON AIRCRAFT FLIGHT SIMULATORS

Table 3-1

2.3. Control points and A/G messages

For convenience, the flight within the integrated control area will be considered as being made up of three phases, as follows.

A preparatory phase, starting when the aircraft is notified to the ZOC Management Centre. It consists in making a prediction of the runway load based on all the information available, estimates of all entries included, and accordingly assigning, tentatively, a preliminary landing slot to the aircraft entering. It ends with the transfer of control and the ground-to-air messages acknowledging the information received prior to entry.

The en route phase, which follows. The first ground-to-air message confirms, or otherwise, cruise altitude and route, then the first ZOC message is prepared and sent to the aircraft. It is sent as early as possible after entry, and includes route, new cruise speed and the position at which to initiate transition from entry speed to new cruise speed, this position being expressed in terms of DME distance from an appropriately located station, Dortmund in the case of the configuration illustrated (Figure 2-1). A second ZOC message will be issued subsequently. This message contains the en route descent characteristics, namely, the position of the initiation of the transition from cruise to descent, noted CTL-2 in Figure 2-1, also expressed in terms of DME distance from a suitably located station (Frankfurt in the illustration), the en-route descent speed profile (Mach/CAS), confirmation of the route, possibly high-altitude holding directives and the estimated arrival conditions (time and altitude) over the next active control point (noted CTL-3 in Figure 2-1) also defined in terms of DME distance from a suitable station, Frankfurt FFH in the example of Figure 2-1.

The final descent phase, including a number of messages and active control points (CTL-3(a), (b), etc.), which will depend on the local geographical configuration and type of approach considered. This phase is critical and will be discussed in detail in a subsequent section.

The tests made to assess the feasibility of the procedure, the accuracy of the 4-d controlled trajectory, during both the en-route and final descent phases, and the range available for correction, whatever the sources of dispersion, will be summarised and discussed in subsequent sections.

3. EN-ROUTE CRUISE/DESCENT PHASE

3.1. Overview

This phase has been discussed previously (Refs. 13 to 16) and it will accordingly be only outlined in the present paper. A total of 32 flights were conducted on the route Pampus (PAM), Dortmund (DOM), Gerdinghausen (GMH), Limburg (LIM), Metro (MTR). The flights were conducted using aircraft flight simulators: Boeing 737 and Airbus A-300 from the Deutsche Lufthansa, Flight Simulation Department, McDonnell Douglas DC-10 from SABENA, Belgian World Airlines and Fokker FK-28 from the Dutch National Aerospace Research Laboratory. A summary of the flights conducted is given in Table 3-1. The control directives to be sent to the pilot were generated on-line. Both voice and automated digital communications were considered, and in each mode of operation the directions were sent some 30 to 60 seconds, depending on the occasion, before the required action had to be initiated.

The aims of the tests conducted covered three essential aspects, namely (a) the practical operational character of the procedure (generation, transmission and acknowledgement of the control directives), (b) the accuracy of the en-route component of the 4-d trajectory from entry into the ZOC control area to the assembly point represented by Metro, 5,000 ft, in these exercises, and (c) the range of control available prior to the final descent.

The scenario of each flight is in line with the procedure described in Section 1. The time of transit is determined at entry (at the latest some 30 to 60 seconds before CTL-1); the relevant CTL-1 position and cruise speed are accordingly sent to the aircraft. Subsequently, the en-route descent characteristics (position of the relevant control point, CTL-2, and speed profile) are also sent from 30 to 60 seconds in advance, and the aircraft proceeds to the assembly point, here Metro. There, the altitude and time are compared against the initial prediction. The procedure to be followed was explained to the pilots prior to their flights by means of a short briefing.

| FLIGHTS | ROUTE | LENGTH | AIRCRAFT | OPERATORS | PERIOD | REPORT |
|---------|---------|--------|----------|-------------------------|---------|---------|
| 5 | RKN-MTR | 140 | B737 | LUFTHANSA PILOTS | JUL. 81 | 812020 |
| 3 | RKN-MTR | 140 | A-300 | LUFTHANSA PILOTS | JUL. 81 | 812020 |
| 7 | PAM-MTR | 201 | DC-10 | SABENA CHIEF INSTRUCTOR | MAR. 82 | 822028 |
| 9 | PAM-MTR | 201 | DC-10 | SABENA CHIEF INSTRUCTOR | NOV. 82 | 832028 |
| 4 | PAM-MTR | 201 | B-737 | LUFTHANSA PILOTS | MAR. 83 | 832028 |
| 2 | PAM-MTR | 201 | A-300 | LUFTHANSA PILOTS | MAR. 83 | 832028 |
| 1 | PAM-MTR | 201 | FK-28 | NLR PILOT | DEC. 83 | 84E1/TN |
| 1 | GMH-MTR | 68 | FK-28 | NLR PILOT | DEC. 83 | 84E1/TN |

SUMMARY OF 32 EN-ROUTE CRUISE/DESCENT TESTS CONDUCTED ON AIRCRAFT FLIGHT SIMULATORS

Table 3-1

3.2. Summary of results

Clearly, the results obtained over a sample of 32 flights cannot claim to have any statistical value. Nevertheless, they provide an amount of information sufficient to establish sound recommendations for subsequent development.

It appears that: (a) the procedure proposed is compatible with present ATC and A/C modes of operation, although significant differences are noticeable when referred to present day operation. These relate, in particular, to the use of combined cruise/descent speed profiles appreciably different from those currently followed. Further, the initiation of the transition from entry to cruise (CTL-1) and from cruise to descent (CTL-2) are determined accurately in terms of DME-related position, which differs appreciably from present practice. Obviously, such points will require particular attention from the crew, besides which, the flights could be considered as "routine", whether conducted manually or using the autopilot modes available.

(b) The trajectory control accuracy has been determined throughout the flight, in particular at the characteristic points (way points and control points). In this summary, it will be sufficient to quote the errors noted at the assembly point, Metro in terms of both altitude and time. Over the set of 32 flights conducted, the absolute value of the error observed in the time of arrival was less than 25 seconds in 84 percent of the cases. The greatest error observed was 44 seconds. The average error (absolute value) was 16 seconds (standard deviation: 11 seconds). The difference in altitude was, in particular, affected by the accurate indication of the cruise to descent transition. If two flights are discarded, simply because it was clear that the pilot had "forgotten" to initiate the descent at the recommended position, the mean value of the error observed (absolute value) is of the order of 260 feet.

3.3. Updates and trajectory corrections

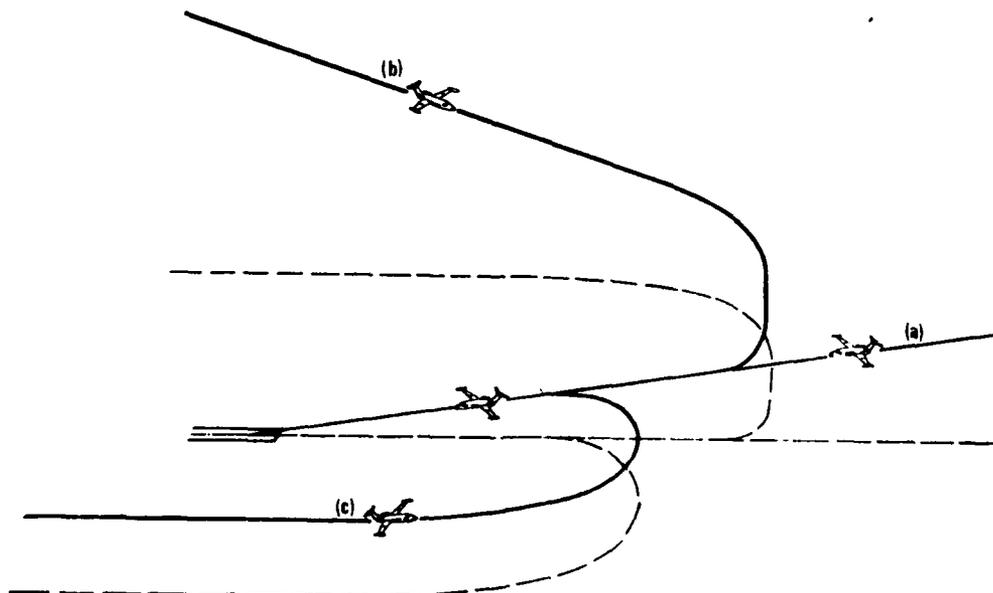
In practice, trajectory updates will be combined with ATC time-of-arrival modifications. During these experiments, ATC corrections were introduced such as to use the entire time-of-arrival control range compatible with the aircraft's operational allowable speed range.

Accordingly, it is expected that the time of arrival over the assembly point can be controlled to within some 30 seconds, this error being accounted for mainly by wind uncertainties during the en-route descent. Now, will it be possible to compensate and provide the necessary corrections to maintain the arrival sequence on the runway? This point is discussed in the next section.

4. FINAL DESCENT AND LANDING

4.1 Terms of the problem

When an aircraft is approaching the final phase of the en-route descent, say over a given geographical point at an altitude of some 10,000 to 5,000 feet, the deviation of the aircraft trajectory from the expected path results from the accumulation of errors of all sources over the descent and possibly part of the cruise.



SCHEMATIC DIRECT AND DOWN-WIND APPROACHES

Figure 4-1

The problem now amounts to amending the subsequent flight phase to ensure arrival on the runway either (a) at the initially requested time or (b) at an earlier or later time to include possible alteration of the landing time sequence, as may be necessary in the light of the developments in the traffic situation.

Several types of approach can be envisaged, as illustrated in Figure 4-1, where (a) corresponds to a direct straight line approach, cases (b) and (c) to down-wind or more generally U-shape approaches, relating to a continuous descent or an interception of the glide path in level flight.

For a number of reasons, it was decided to conduct the flights in the Brussels area, and two basic types of approach were considered, namely a direct straight line approach and a typical U-shape approach of which the essential characteristics will be defined in Section 4.2.

Some 21 flights were then conducted using the SABENA McDonnell Douglas DC-10 aircraft flight simulator. The main objectives included the confirmation of the range of control available, the feasibility of controlling 4-d flights with a high degree of accuracy in present modes of operation, and the effect of essential control parameters on the overall performance.

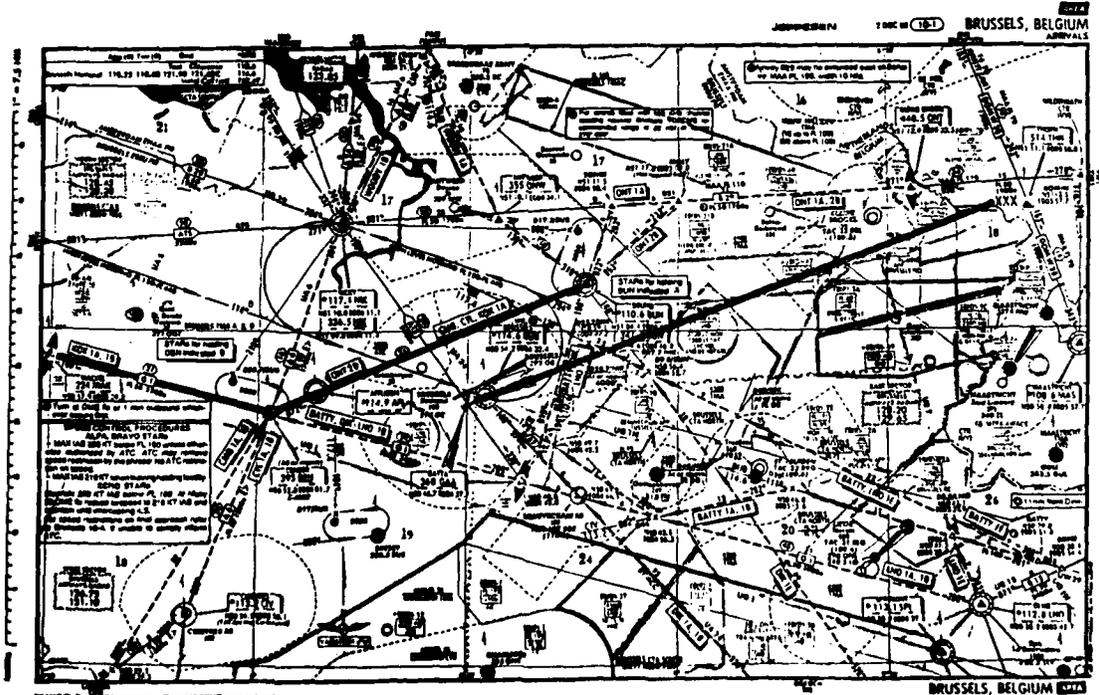
4.2 Straight line and U-shape approaches

At the time of the exercise, the arrival characteristics in the Brussels area appeared as shown in the Map (Jeppesen, Arrivals, Brussels Area, 2 Dec. 1983) reproduced in Figure 4-2 (Courtesy of Jeppesen GmbH, Frankfurt, FRG). The two approach routes considered are drawn in thick lines. The straight line approach has been conceived as follows: It extends over 50 nm from XXX, where the flights were initiated, to the runway (25 L). The U-shape approaches are conducted along the existing route Mackel-Dender-Bruno, the initiation of the turning manoeuvre taking place at a precise distance from Bruno, as discussed subsequently.

4.2.1 Control of a straight line approach

The control of a straight line approach, as it is considered here, will be described using the information given in Figure 4-3. In the upper part of the diagram, the route is shown in dotted lines (x and y coordinates (eastings and northings, respectively) are referred to an origin chosen on the runway). The vertical distance between the geographical route and the continuous curve is proportional to the altitude, providing a three-dimensional view of the flight geometry. The lower part of the diagram gives the altitude and indicated airspeed as recorded during the flight.

The aircraft arrives over XXX at an altitude of 10,000 ft at a speed of 250 kt (IAS). The relevant ZOC message is generated and sent to the pilot. It contains the position of Control Point 3 (CTL-3), here 39nm from BUB (DME-related distance) and the definition of the action, here a deceleration to 233 kt (IAS), then the descent at that speed to 2,000 feet and, at that altitude, interception of the ILS glide path.



ARRIVAL ROUTES SELECTED FOR THE EXERCISE

(a) XXX-BUB (added to existing chart) (b) MAE-DEN-(= BRU)-BUB

Chart reduced from Brussels 10-1 Area Chart, 2 Dec. 83
Copyright 19, 8 August 1984, Jeppesen S Co

Figure 4-2

In such an approach type, the en-route descent speed is the only practical control of the arrival time. With flights starting from 10,000 ft at 250 kt (IAS), the range of control extends over about 3 minutes (for delays only) as Table 4-1 indicates. Clearly, reducing the speed in this phase will increase the consumption and a fortiori the cost. This increase may be appreciable. In the case of a DC-10, it is of the order of 150 kg per minute of delay.

4.2.2 U-shape approach

Figure 4-4 shows a U-shape approach and its essential control characteristics (the same type of presentation as was used for the straight line approach). In this illustration, the aircraft comes from Mackel, it is in level flight, at altitude 10,000 ft and a speed of 250 kt (IAS). The procedure includes three control points (CTL-3A to C), where the relevant directives contained in the corresponding ZOC messages have to be initiated. As for the en-route parts, the messages are available at least 30 to 90 seconds before the times of initiation. Position and CTL-3A and B are defined in terms of BUN DME-related distances, while the position of CTL-3C is determined by the orientation of the axis aircraft - Brussels outer marker beacon(OB).

The directives (contents of ZOC messages 3A to C) corresponding to a particular flight are listed in the upper part of the diagram (Fig. 4-4).

In such an approach type, the control possibilities include, of course, the en-route descent speed adjustment as for a straight line approach. Nevertheless, in addition, the parameters which govern the final path provide further control for the time of arrival: some 2.5 minutes. This will be discussed in more detail below.

4.3 Control of time of arrival: Range and governing parameters

4.3.1 Straight line approach

As already indicated, the governing control parameter is the descent indicated airspeed. It has been pointed out above that when the adjustment (correction for previously accumulated errors and/or modification of the landing time slot as a result of traffic evolution) is initiated at 10,000 feet, the range available amounts to three minutes. Clearly, this range increases if the ZOC control point CTL-3 is envisaged at a higher altitude. Table 4-1 summarizes the results obtained for a series of five flights, the initial en-route descent being conducted at 250 kt (IAS), and the adjustments resulting from decelerations to 233, 210, 185 and 170 kt (IAS), respectively. An illustration of the flights observed as a result of the directives (such as shown in Figure 4-3, upper part) is given in Figure 4-5 for the horizontal view, altitude and speed versus distance.

4.3.2 U-shape approach

Besides the descent speed, which of course plays a role similar to that in the straight line approach, additional governing parameters are available to adjust the time of arrival on the runway, namely the position of the turning point to the baseleg (CTL-3B), the direction of this segment (CTL-3C) and possibly, the top-of-descent or transition to descent position (CTL-3A) where a phase of horizontal flight has been introduced at low altitude (such as 10,000 ft in Figure 4-4).

The range of control associated with each of these parameters appears in Table 4-2, while Figure 4-6 illustrates the use of descent speed control. It shows two flights following the same approach (that of flight F1 in Table 4-2 (a)), the definition of the flights differing in the descent speeds from 10,000 feet, namely 250 and 185 kt. The arrival times differ by 1.3 minutes, as against 2.3 minutes for the straight line approach. When the position of the turning point (CTL-3B) varies from 6.1 to 2.1 nm, DME-related distance from Bruno, the time of arrival increases by some 2.3 minutes. The flights are given in Table 4-2(a), and an illustration (horizontal phase) is shown in Figure 4-7. If the heading of the baseleg then varies from 184 to 144 deg., other parameters being kept unchanged - see Table 4-2(b) and Figure 4-8 - the time of arrival varies by some 43 seconds. A modification of the position of control point CTL-3A, namely top-of-descent from 10,000 ft, see Figure 4-4, of 7nm, DME-related distance from BUN, would provide the same variation but at a higher cost (see Table 4-2 (c)). An illustration of the flights made (F6, F7, F8) is given in Figure 4-9.

Legend for Tables 4-1 to 4-4

- d : DME-related position, in nm (BUB for straight line approaches; BUN for U-shape approaches)
(TOD: start of final descent; bl: position at which to initiate turn to base-leg)
- m_f : fuel consumed, in kg
- m_{a1} : aircraft initial mass, in tons
- t : arrival time over the runway, in min. (ap: predicted; a: observed)
- v_{de} : descent speed, in kt (IAS)
- θ_{bl} : base-leg heading, in deg.
- θ_{ap} : heading to outer marker beacon at which to initiate turn to glide path, in deg.
- τ_A : difference between observed and requested time of arrival at the runway, in sec.
- τ_{AR} : difference in time-of-arrival for repeated flights, in sec.

| Flight No | CTL-2 (nm) | v _{de} (kt) | t _{ap} (min) | ' _a (sec) | m _f (kg) |
|-----------|------------|----------------------|-----------------------|----------------------|---------------------|
| F20 | 39 | 250 | 11.37 | + 4 | 836 |
| F16 | 39 | 233 | 11.52 | +11 | 796 |
| F17 | 39 | 210 | 12.35 | +21 | 998 |
| F18 | 39 | 185 | 13.68 | + 2 | 1175 |
| F19 | 39 | 170 | 14.35 | +13 | 1240 |

STRAIGHT LINE APPROACH
INFLUENCE OF EN-ROUTE DESCENT SPEED

Table 4-1

a) Effect of position of turning point to base-leg

| Flight No | d _{TOD} (nm) | v _{de} (kt) | d _{bl} (nm) | θ _{bl} (deg) | γ _{gp} (deg) | t _{ap} (min) | ' _a (sec) | m _f (kg) |
|-----------|-----------------------|----------------------|----------------------|-----------------------|-----------------------|-----------------------|----------------------|---------------------|
| F1 | 30.6 | 250 | 6.1 | 164 | 234 | 14.87 | +3 | 1285 |
| F2 | 30.6 | 250 | 5.1 | 164 | 237 | 15.45 | +5 | 1382 |
| F3 | 30.6 | 250 | 4.1 | 164 | 239 | 15.94 | 0 | 1422 |
| F4 | 30.6 | 250 | 3.1 | 164 | 240 | 16.52 | +3 | 1448 |
| F6 | 30.6 | 250 | 2.1 | 164 | 241 | 17.19 | +9 | 1573 |

b) Effect of base-leg heading

| Flight No | d _{TOD} (nm) | v _{de} (kt) | d _{bl} (nm) | θ _{bl} (deg) | γ _{gp} (deg) | t _{ap} (min) | ' _a (sec) | m _f (kg) |
|-----------|-----------------------|----------------------|----------------------|-----------------------|-----------------------|-----------------------|----------------------|---------------------|
| F13 | 24.4 | 250 | 3.8 | 184 | 235 | 15.88 | +8 | 1353 |
| F14 | 24.4 | 250 | 3.8 | 164 | 237 | 15.42 | +14 | 1337 |
| F15 | 24.4 | 250 | 3.8 | 144 | 240 | 15.16 | +4 | 1298 |

c) Effect of top-of-descent position

| Flight No | d _{TOD} (nm) | v _{de} (kt) | d _{bl} (nm) | θ _{bl} (deg) | γ _{gp} (deg) | t _{ap} (min) | ' _a (sec) | m _f (kg) |
|-----------|-----------------------|----------------------|----------------------|-----------------------|-----------------------|-----------------------|----------------------|---------------------|
| F6 | 30.6 | 250 | 2.1 | 164 | 241 | 17.19 | +9 | 1573 |
| F7 | 26.6 | 250 | 2.1 | 164 | 241 | 16.77 | -8 | 1455 |
| F8 | 23.6 | 250 | 2.1 | 164 | 241 | 16.46 | +5 | 1445 |

U-SHAPE APPROACHES
INFLUENCE OF MAIN CONTROL VARIABLES

Table 4-2

4.4 Accuracy of time of arrival control

4.4.1 Observation versus prediction

The time of arrival was predicted from the initial point (XXX and Mackel for straight line and U-shape approaches respectively, in both cases initial altitude 10,000 ft, initial speed 250 kt IAS). No subsequent update was made, which obviously would be possible in practice, in particular to account for possible subsequent deviations.

The results obtained are summarized in Tables 4-1 and 4-2. The predicted time of arrival is noted t_{ap} , and the difference between the time of touch-down and the predicted value is noted ϵ_a . It is positive when the aircraft lands later than predicted.

During a U-shape approach the maximum error observed did not exceed 14 seconds for a series of 15 flights. During the experiments, six straight line approaches were made, and the maximum discrepancy reached 21 seconds. For the two types of approach the average values of the absolute errors were found to be 6 and 9 seconds respectively.

4.4.2 Influence of initial mass on prediction accuracy

All the calculations used for predictions were made on the assumption of an aircraft of initial mass of 160 tons. Various investigations had indicated that for en-route descent the effect of mass could probably be neglected and incorporated in the adjustments. To confirm or otherwise, the initial mass of the aircraft was taken as being equal to 180 tons and the predictions made on the assumption of a 160-ton aircraft. The results obtained are presented in Table 4-3 and the flights illustrated in Figures 4-10 and 4-11. For the straight line approach, the error (8 sec) is within the order of magnitude noted previously (Table 4-1). For the U-shape approach it is of the same order (17 sec) as the highest value (14 sec) observed during the series of 15 flights for which the predictions were made using the actual aircraft mass.

| Flight No | m_{ai} (tons) | t_{ap} (min) | ϵ_a (sec) | m_f (kg) | Flight No | m_{ai} (tons) | t_{ap} (min) | ϵ_a (sec) | m_f (kg) |
|-----------|-----------------|----------------|--------------------|------------|-----------|-----------------|----------------|--------------------|------------|
| F1 | 160 | 14.87 | + 3 | 1285 | F18 | 160 | 13.68 | + 2 | 1175 |
| F11 | 180 | 14.98 | +17 | 1425 | F21 | 180 | 13.68 | + 8 | 1275 |

a) U-shape approach (d_{TOD}/d_{bl} : 30.6/6.1 nm from BUN) b) Straight line approach (v_{de} = 185 kt)

INFLUENCE OF INITIAL MASS ON PREDICTION ACCURACY

Table 4-3

| Flight No | t_a (min) | ϵ_{ar} (sec) | m_f (kg) | Flight No | t_a (min) | ϵ_{ar} (sec) | m_f (kg) |
|-----------|-------------|-----------------------|------------|-----------|-------------|-----------------------|------------|
| F1 | 14.90 | + 3 | 1285 | F3 | 15.94 | 0 | 1422 |
| F5 | 14.87 | 0 | 1287 | F9 | 16.06 | + 12 | 1425 |
| F12 | 15.01 | +14 | 1295 | | | | |

a) d_{TOD}/d_{bl} : 30.6/4.1 nm from BUN

b) d_{TOD}/d_{bl} : 30.6/6.1 nm from BUN

REPEATABILITY OF FLIGHT CONTROL

OBSERVATIONS MADE DURING U-SHAPE APPROACHES

Table 4-4

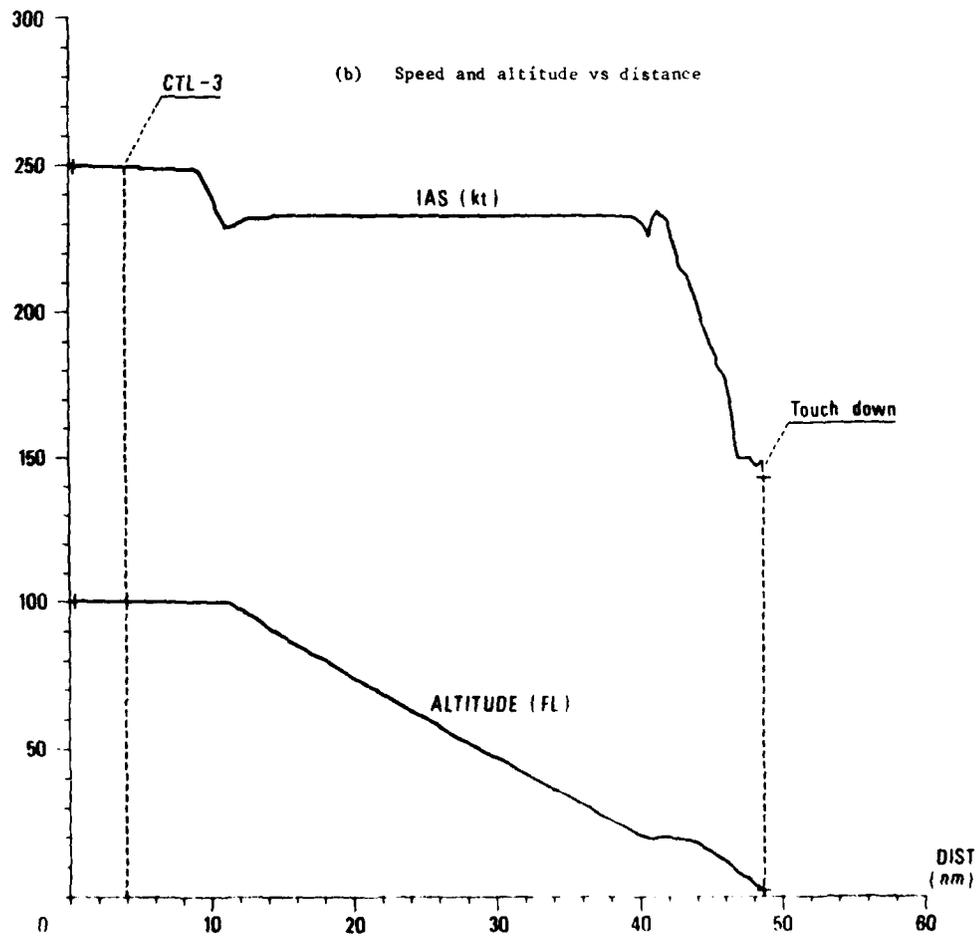
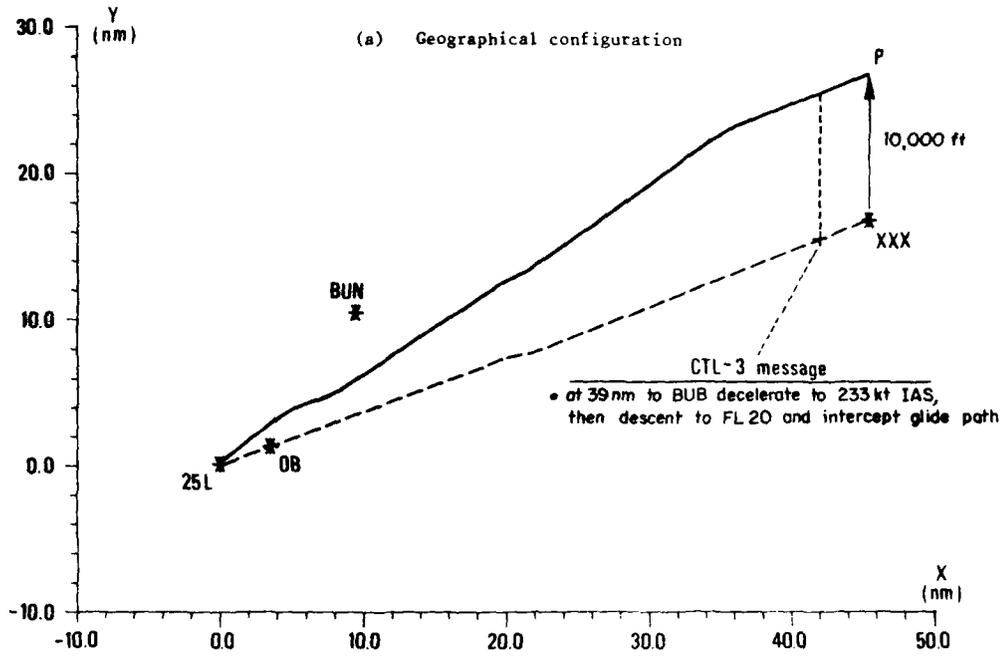


ILLUSTRATION OF A STRAIGHT LINE APPROACH

Figure 4-3

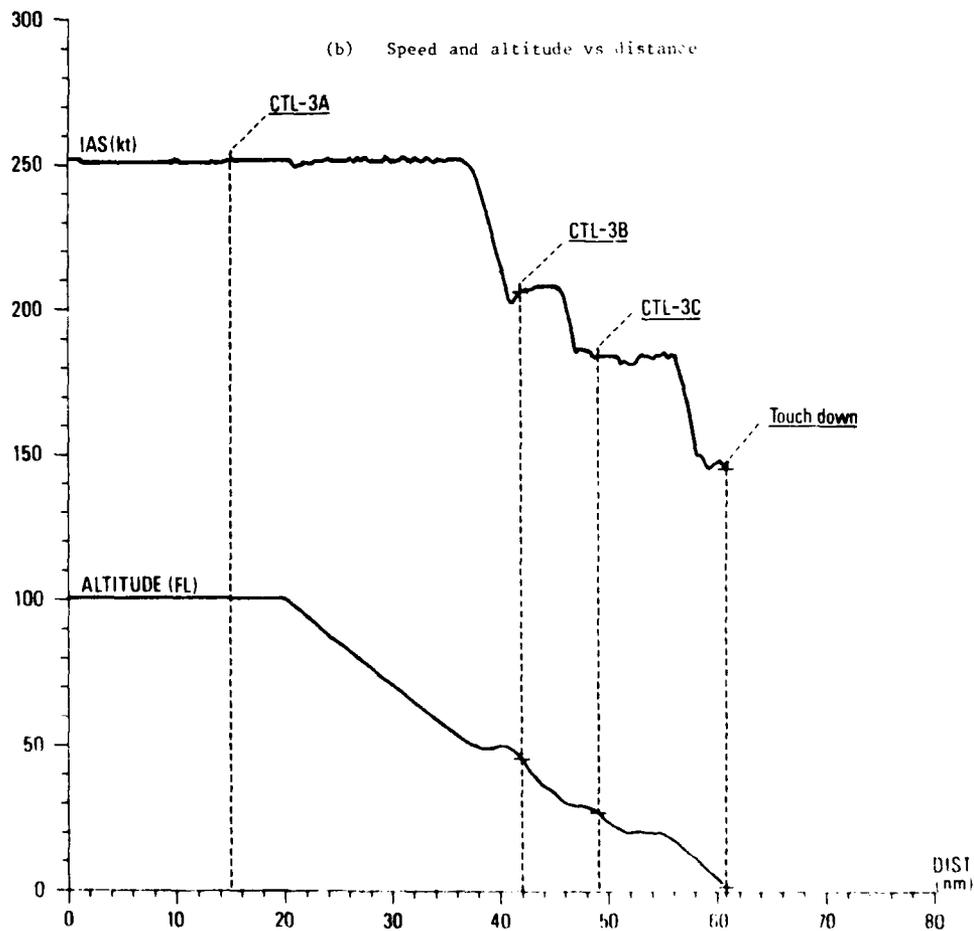
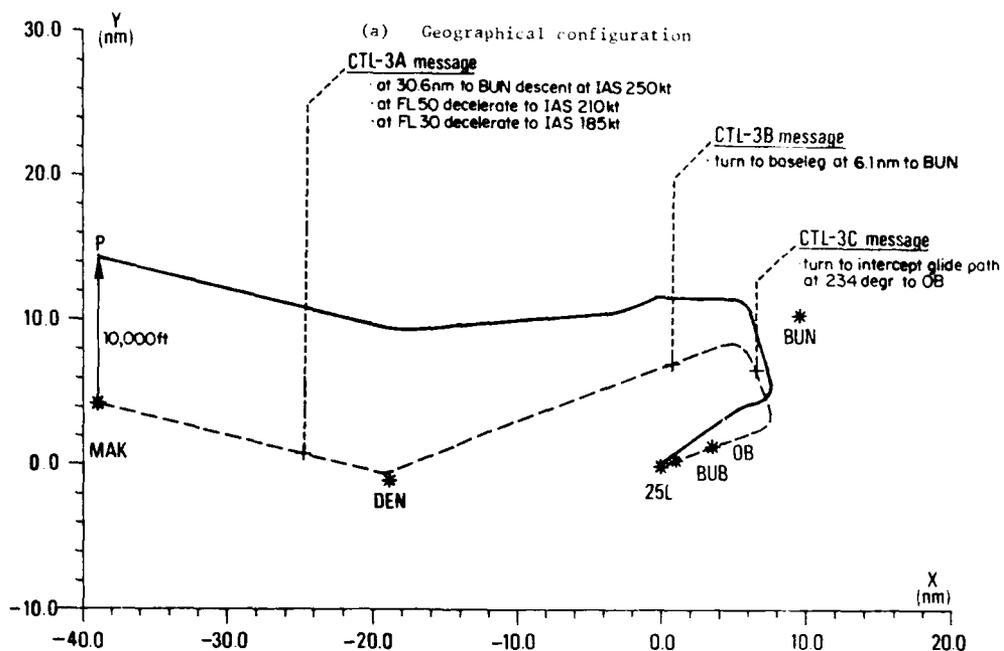
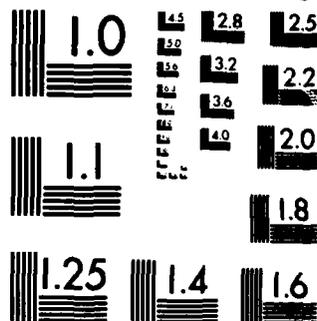
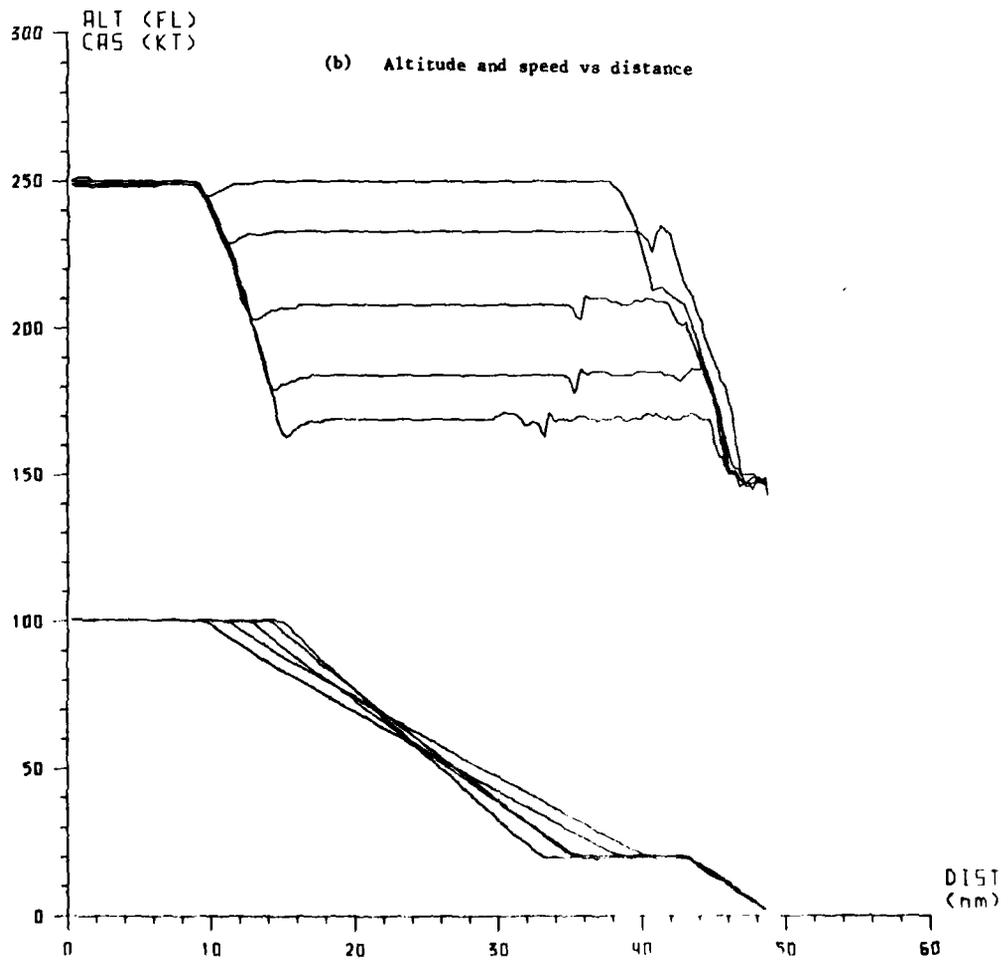
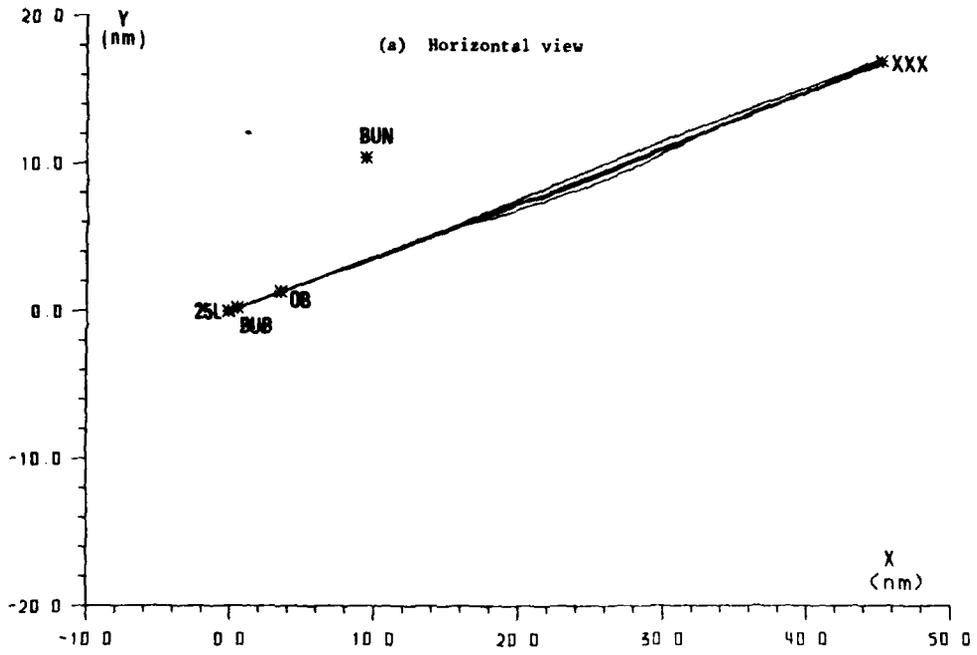


ILLUSTRATION OF A B-SHAPE APPROACH

Figure 3-3

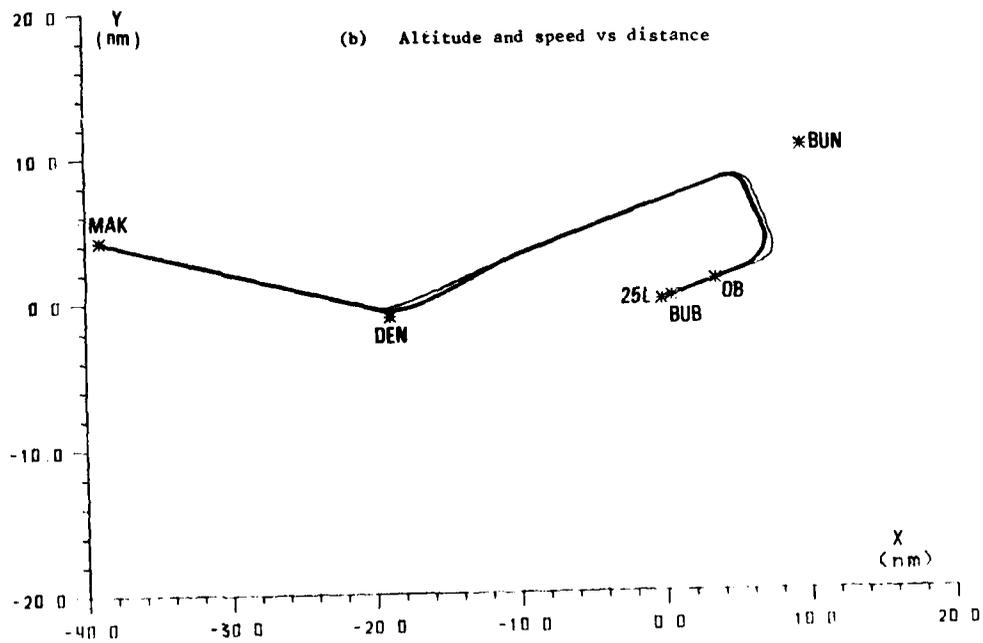
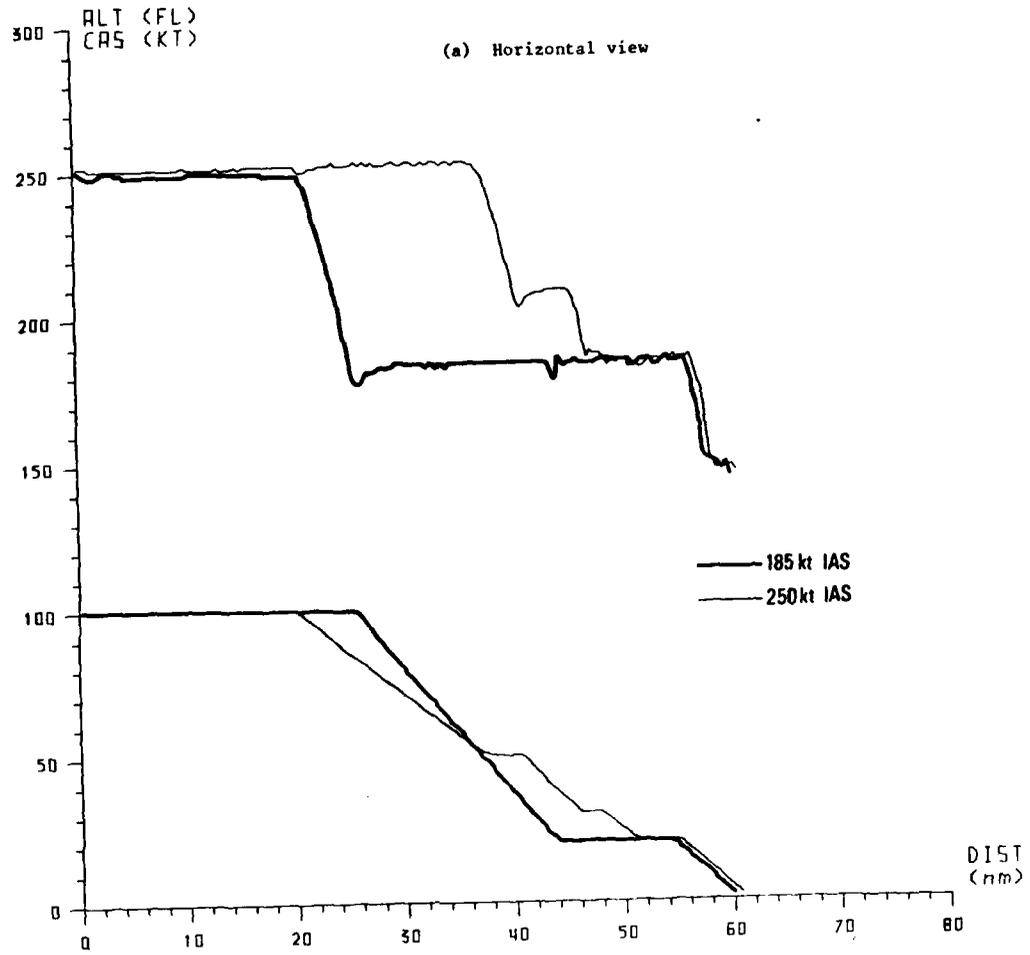


MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A



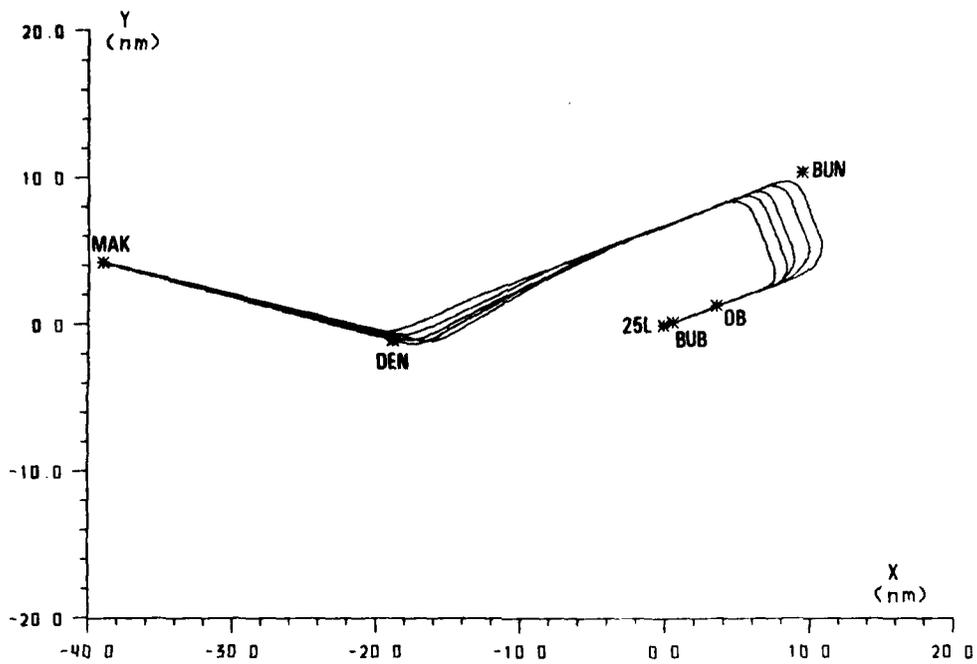
STRAIGHT LINE APPROACH: INFLUENCE OF DESCENT SPEED

Figure 4-5



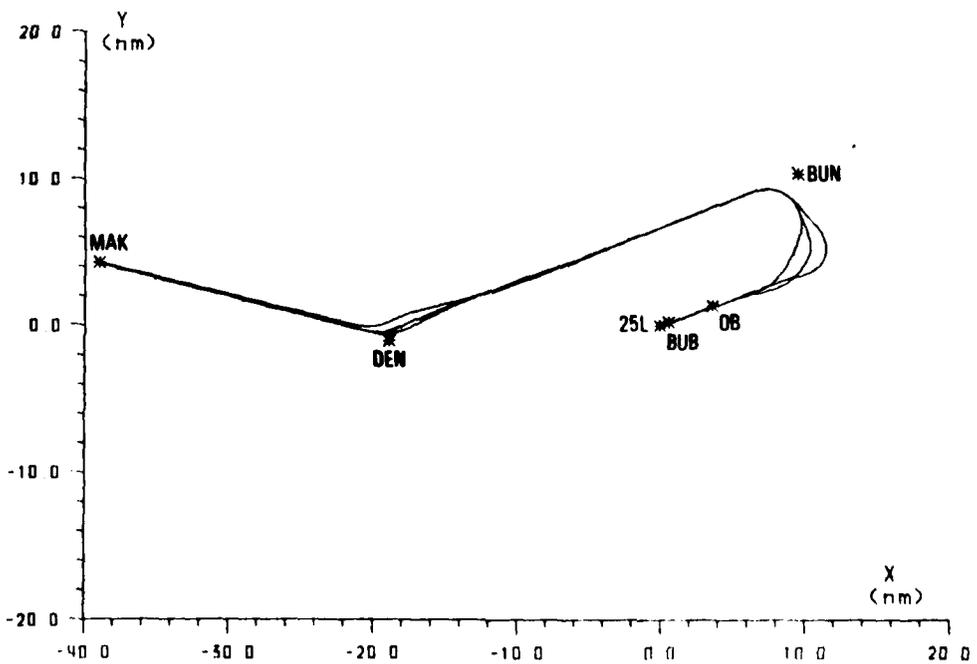
U-SHAPE APPROACH: INFLUENCE OF DESCENT SPEED

Figure 4-6



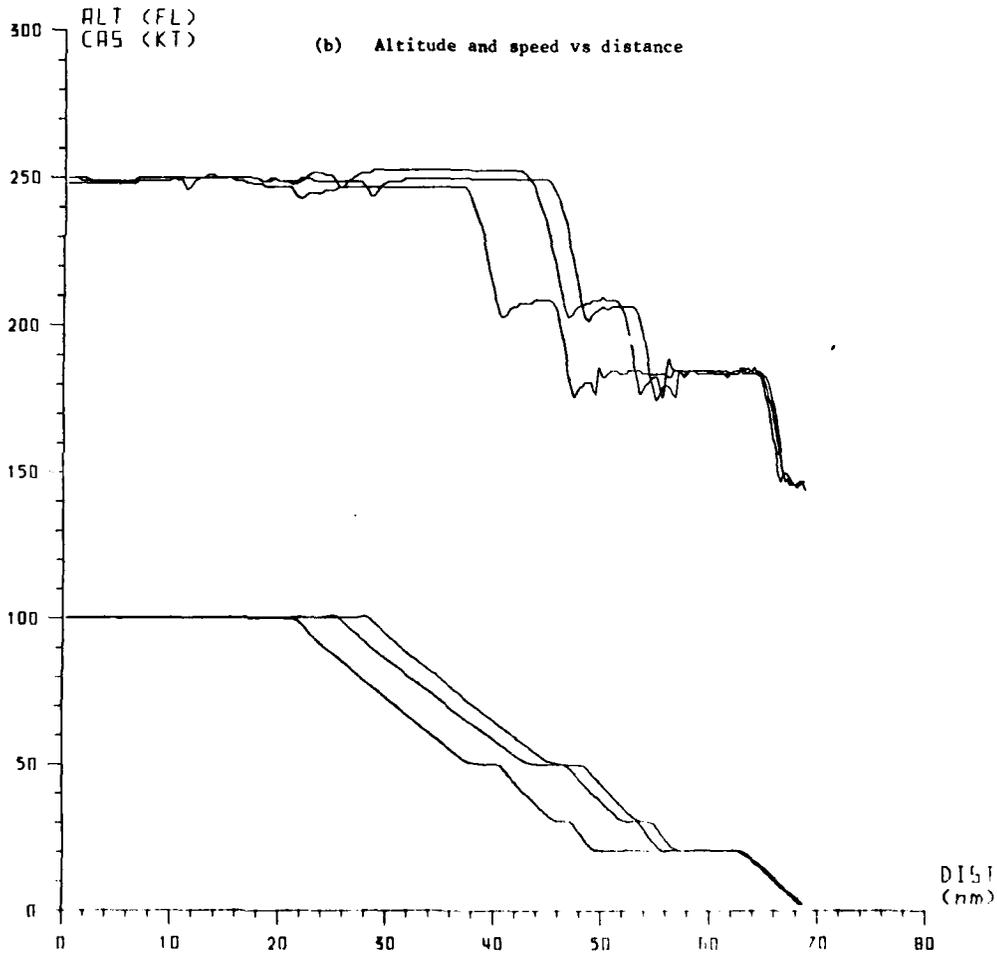
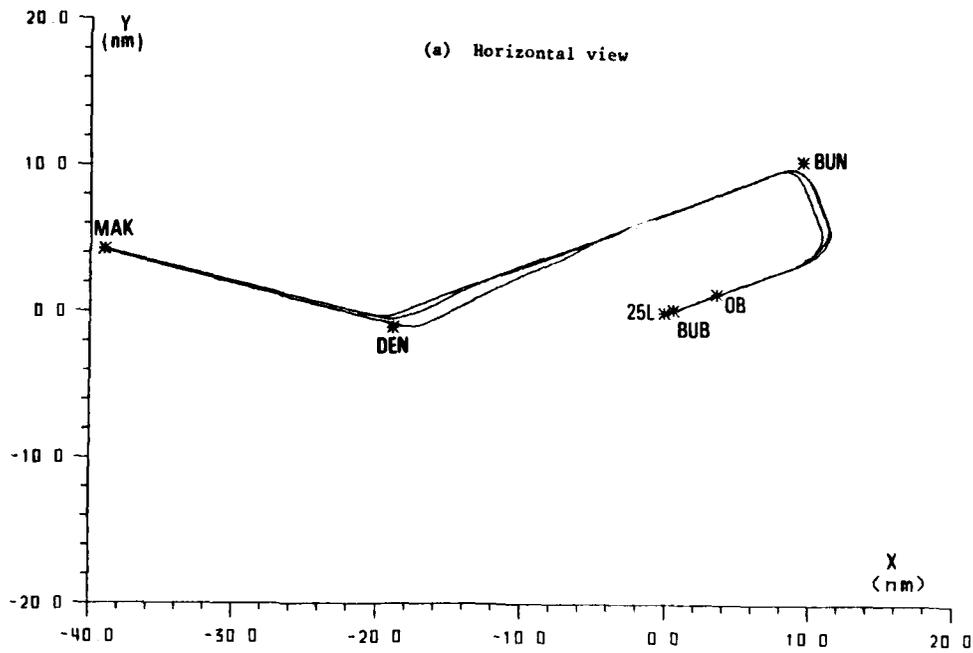
U-SHAPE APPROACH: INFLUENCE OF POSITION OF TURNING POINT TO BASE-LEG

Figure 4-7



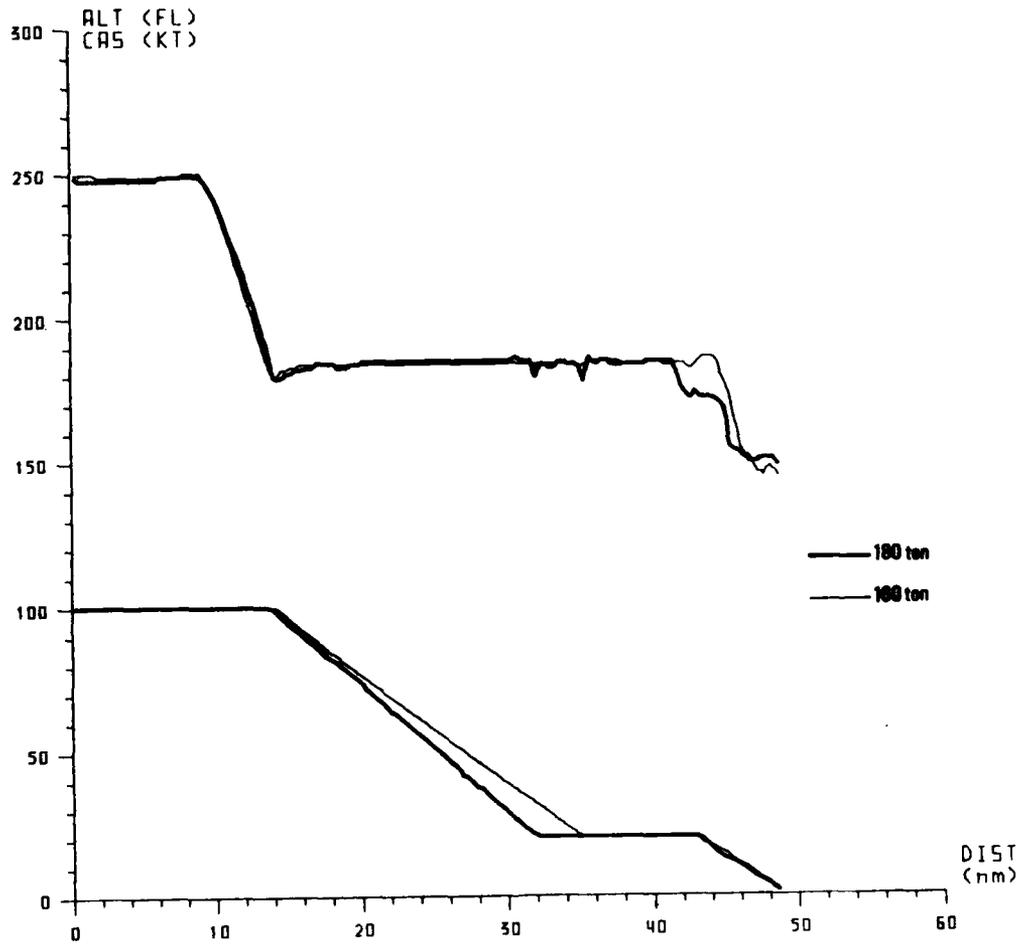
U-SHAPE APPROACH : INFLUENCE OF BASE-LEG HEADING

Figure 4-8



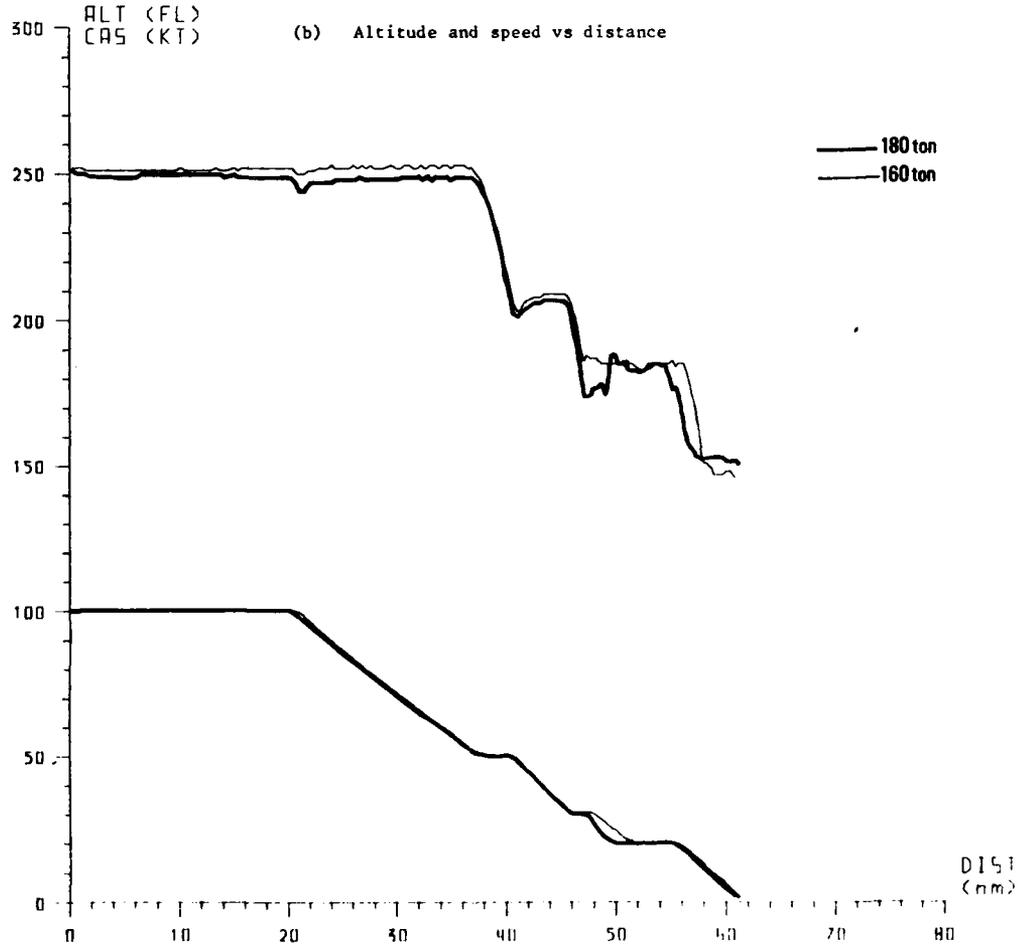
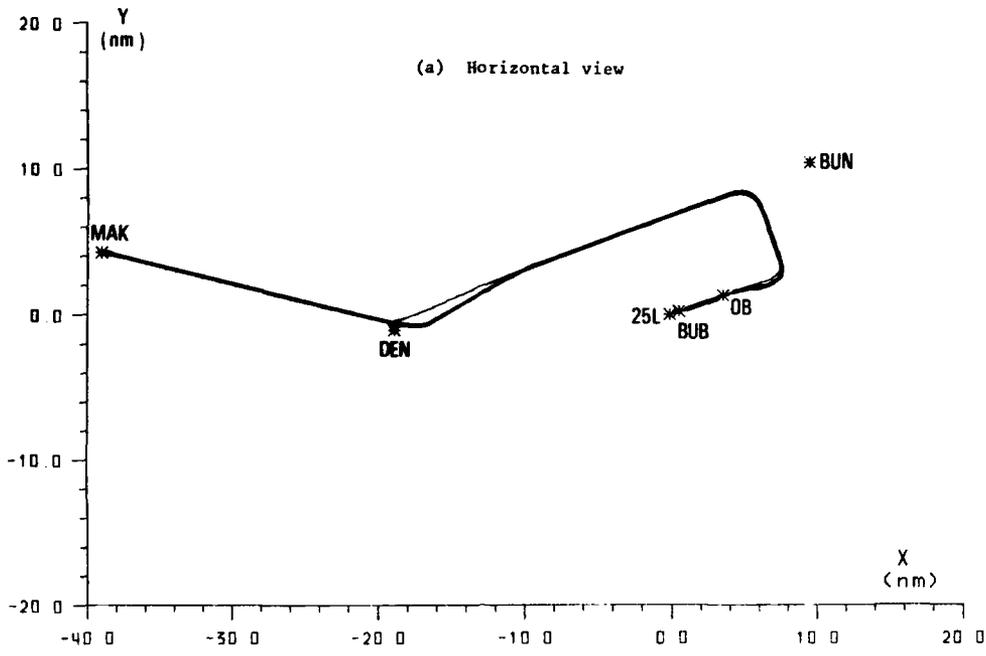
U-SHAPE APPROACH : INFLUENCE OF POSITION OF START OF FINAL DESCENT

Figure 4-9



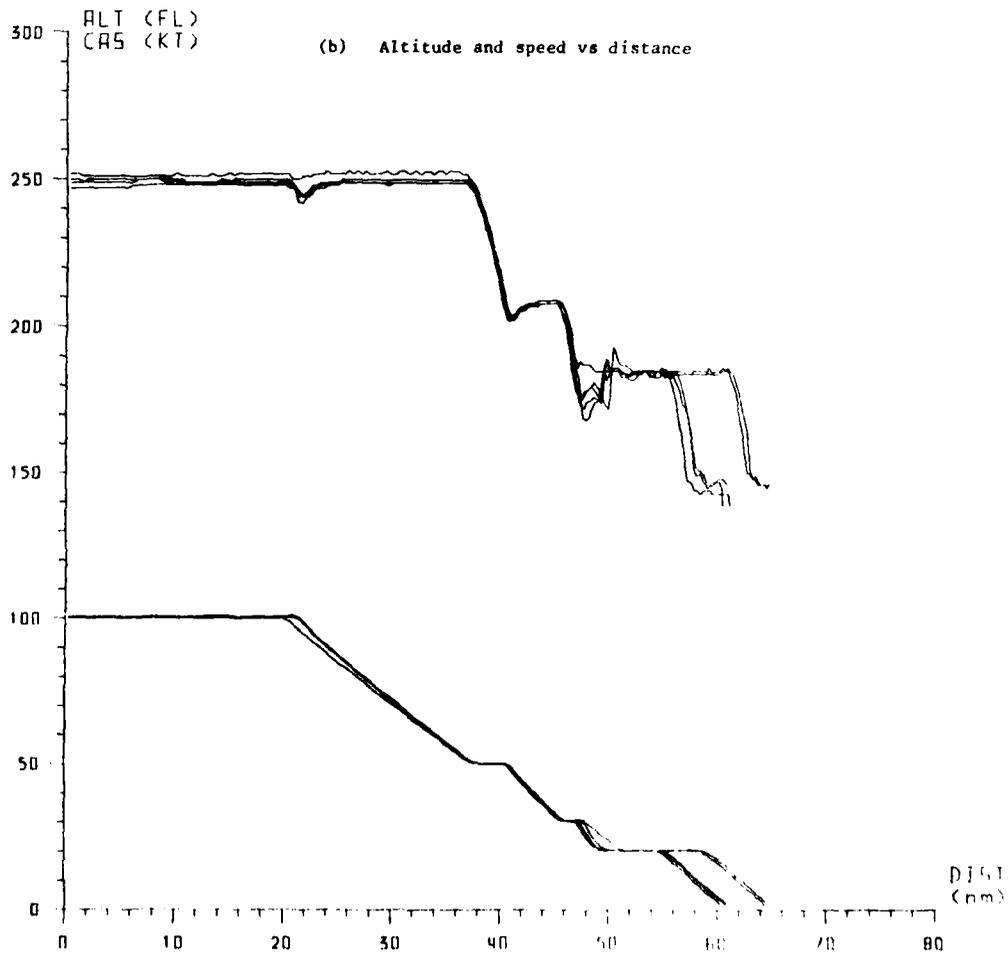
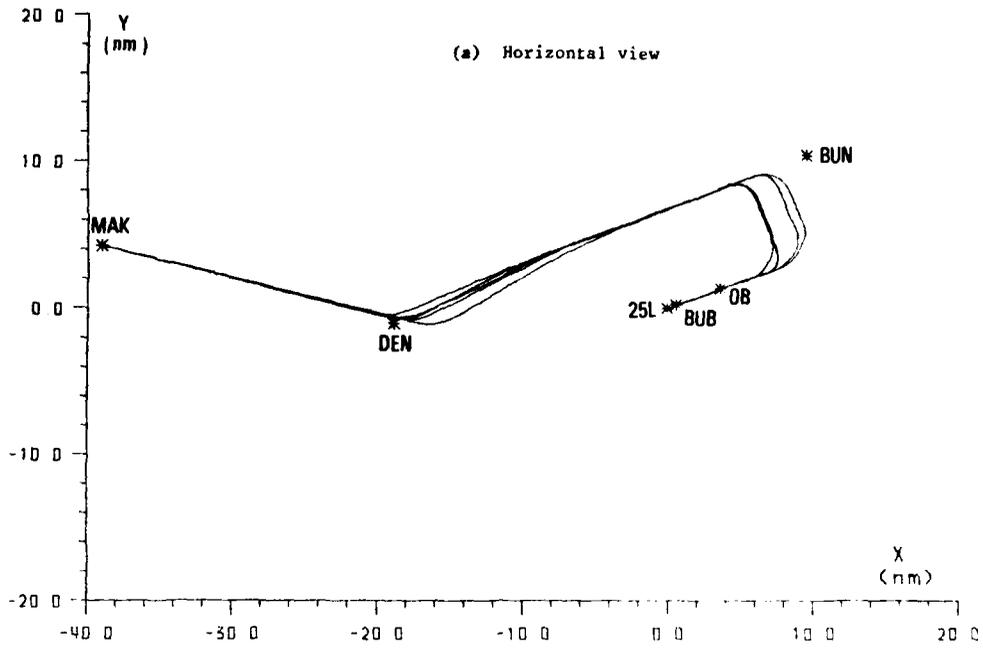
STRAIGHT LINE APPROACH : EFFECT OF AIRCRAFT MASS UNCERTAINTY

Figure 4-10



U-SHAPE APPROACH : EFFECT OF AIRCRAFT MASS UNCERTAINTY

Figure 4-11



U-SHAPE APPROACH: REPEATABILITY OF FLIGHTS

Figure 4-12

4.4.3 Repeatability of accurately controlled approaches

Clearly, it was of interest, (a) to see whether the conduct of flights defined in an identical manner (same set of CTL-3 messages, in particular in a U-shaped approach) would be conducted in the same way and lead to the same results, and (b) to assess the dispersion observed. For this purpose, two flights were repeated, once or twice, two days later. These are flights F1 and F3, defined in Table 4.2.

The results are presented in Table 4-4 and illustrated in Figure 4-12. The repeatability is good, and the discrepancies observed (3, 12, 14 seconds) are within the range of predictability errors noted for the entire set of flights.

5. CONCLUSIONS

Present air traffic developments, together with economic constraints, call for an efficient use of the capacity available, in particular landing capacity, and a drastic reduction in flight cost wherever possible. For inbound traffic, this in turn requires effective cooperation between en-route and approach control units (a) to determine the optimum set of landing times and (b) to control each individual flight accordingly.

A specific procedure is proposed for the accurate control of individual 4-d trajectories, time-of-arrival-constrained, over an extended distance (up to 250nm from touch-down) using the guidance, control and navigation equipment currently available. To this effect, the flight is considered as being made up of two main phases, each being controlled by specific variables:

- (a) The en-route phase, which includes the cruise and the en-route descent; it is essentially controlled through the combined cruise/descent speed profile;
- (b) The final descent, which covers the transition from en-route descent, approach part, ILS-glide slope intercept and landing.

The active en-route control phase starts when the aircraft enters the integrated control area (such as a ZOC area extending over some 140 to 250 nm) and is assigned a tentative landing time slot. The relevant message including the cruise speed component is sent to the pilot with the precise position - expressed in DME-related distance from a suitably located station - at which to initiate the transition from entry speed to required cruise speed. Subsequently, the ATC control system - in view of the evolution of the traffic situation - may either confirm or alter the initially allocated landing slot. In parallel to the traffic evolution, the progress of this particular flight may differ from what was expected and an update may become necessary. Whether or not the time of arrival on the runway - or at the assembly point - is confirmed or modified, the possible related adjustment will include any necessary trajectory correction. This leads to a directive defining the en-route descent speed profile. The position at which to initiate the transition from cruise to descent is given to the pilot in terms of DME-related distance (also from a suitably located station), the relevant message being sent from 30 to 60 seconds beforehand. Accordingly, at the assembly point, say between 15,000 and 5,000 ft, the errors accumulated should result only from the en-route descent.

The final descent may be initiated at any altitude, say between 15,000 and 5,000 ft, depending on the magnitude of the adjustment of the time of arrival on the runway which may be required. This phase may or may not include a horizontal part, it may lead to a direct straight line approach or to a particular geometrical shape matching down-wind requirements and local geographical configurations. In the case of a straight line approach, the en-route descent speed is the only practical control variable. For a descent of a DC-10 decelerated to 250 kt (IAS) at 10,000 ft, the speed range available makes it possible to adjust the arrival time on the runway within some three minutes. For a U-shape approach (from which an S-shape could readily be derived whenever necessary) the control variables include, in addition to the route descent speed as for the straight line approach, the position at which to resume the descent if interrupted at, say, 10,000 ft or to initiate the en-route descent, the position at which to initiate the turn to the base-leg and the direction of the base-leg which governs the angle of interception with the glide path. Positions for initiating (see above) or resuming the en-route descent and initiating the turn to the base-leg are given accurately in terms of DME-related distances. The position at which the glide-slope intercept manoeuvre is started is defined by a heading to the outer marker beacon. The adjustments of the time of arrival on the runway made possible through these three parameters (en-route descent speed control not included) depend mainly on the airspace available. In the geographical configuration used during the tests the observed ranges are found to be of the order of 140, 40, and 40 seconds respectively, when considered independently. If these are duly combined, it is felt that adjustments of the order of 2 minutes could readily be considered.

The proposed procedure was exposed to a series of tests - for both the en-route and the final phases - using aircraft flight simulators. The results obtained to date indicate that the following can be achieved.

- (a) compatibility of the procedure with both current aircraft operation and air traffic controllers' directives structure;
- (b) compatibility with present human or near-future synthetic voice communications;
- (c) compatibility with future air/ground automated digital communications;
- (d) suitability to control accurately flights in any ATC system aiming at precise times to arrive at assembly points and/or on the runway;
- (e) possibility of flying 4-d trajectories with current standard airborne equipment, in compatibility with present operation, with a high degree of accuracy. It is felt that control of arrival time to within some 10 seconds over a 200 nm flight distance, should be aimed at.

In conclusion, the procedure discussed in this paper appears suitable to control aircraft in a Zone-of-Convergence-type system making it possible to realise the expected benefits in terms of economy, cost, consumption and efficient use of available capacity.

Further assessment will include a series of tests conducted on-line from an operational ATC centre, and the related developments needed both on the operational side (in particular the form of the A/G messages) and in technical terms (in particular, automatic generation of control directives).

6. ACKNOWLEDGEMENTS

The authors wish to express their appreciation to the members of the EUROCONTROL Specialist Panel on Automatic Conflict Detection and Resolution (SPACDAR) for their critical comments.

The tests reported in this paper were performed on various flight simulators (DC-10, SABENA, Belgian World Airlines; Airbus A-300 and Boeing B-737, Deutsche LUFTHANSA; Fokker F-28, National Lucht-en Ruimtevaart Laboratorium, Amsterdam) operated under contract to EUROCONTROL.

DISCLAIMER

*The views expressed in this paper are those of the authors.
They do not necessarily reflect the policy of the EUROCONTROL Agency.*

REFERENCES

1. Renteux, J.-L. and Schröter, H. "Fuel saving in air transport - Possible contributions of Air Traffic Services in Europe". EUROCONTROL Doc. 812007, February 1982.
2. Barber, D. "Fuel conservation and economy constraints". Paper presented at the AGARD Conference on "Air Traffic Control in Face of Users Demand and Economy Constraints", Lisbon, Portugal, 15 October, 1982. AGARD CP-340, 1983.
3. Benoît, A. and Swierstra, S. "Potential fuel consumption savings in medium- to high-density extended terminal areas". EUROCONTROL Report 832004, March 1983.
4. Imbert, N., Fossard, A.J. et Comes, M. "Gestion à moyen terme du trafic aérien en zone de convergence". IFAC/IPORS Symposium proceedings, Toulouse, France, 6-8 March 1979.
5. Benoît, A. and Swierstra, S. "Air Traffic Control in a Zone of Convergence. Assessment within Belgian Airspace". Congress of the International Council of Aeronautical Sciences, Toulouse, France, 9-14 September, 1984. Also EUROCONTROL Report 842009, April 1984.
6. Völkera, U. "System "COMPAS" für einen rationelleren Anflugverkehr". DFLVLF-Nachrichten, Heft 39, Braunschweig, F.R.G., June 1983.
7. Schwarzott, W.P. and Völkera, U. "The German Concept COMPAS (Computer Oriented Metering Planning and Advisory System)". Paper presented at the EUROCONTROL Specialist Panel on Automatic Conflict Detection and Resolution (SPACDAR), Brussels, September 1983.
8. Ten Have, J.M. and Schrier D. "Computer assisted departure and arrival planning in the Netherlands SARP system". Paper presented at the EUROCONTROL Specialist Panel on Automatic Conflict Detection and Resolution (SPACDAR), Brussels, September 1983.
9. Ord, G. "Concepts for a future system of air traffic management and control in the South-East of the United Kingdom". Paper presented at the EUROCONTROL Specialist Panel on Automatic Conflict Detection and Resolution (SPACDAR), Brussels, 21-23 September 1983.
10. Garnier, J.L. et de Ronne, M. "Résorption économique des délais à l'arrivée". International Seminar on "ATC Contributions to Fuel Economy", EUROCONTROL Institute of Air Navigation Services, Luxembourg, 26-28 October 1982.
11. Adam, V. "Investigations of four-dimensional guidance in the TMA". Paper presented at the AGARD Conference on "Air Traffic Control in Face of Users Demand and Economy Constraints", Lisbon, Portugal, 15 October 1982. AGARD CP-340, 1983.
12. Benoît, A. and Swierstra, S. "A ground/air coordinated control procedure for minimum cost operation over an extended area". Conference on "Today & Tomorrow Mini & Micro Computers in Airline Operations", Royal Aeronautical Society, London, U.K., 21-22 October 1982. Also EUROCONTROL Report 822036, September 1982.
13. Attwooll, V. and Benoît, A. "Fuel economies effected by the use of FMS in an advanced TMA system". Paper presented at the Royal Institute of Navigation, London, U.K., 15 February 1984. Also EUROCONTROL Report 842003, February 1984.
14. Benoît, A. and Swierstra, S. "The control of cruise-descent profiles. Experiments using B-737 and Airbus flight simulators". EUROCONTROL Report 812020, August 1981.
15. Benoît, A. and Swierstra, S. "The dynamic control of inbound flights. Experiments conducted on the SABENA DC-10 flight simulator". EUROCONTROL Report 822028, June 1982.
16. Benoît, A., Garcia, C. and Swierstra, S. "Ground-air coordinated control of time of arrival. Part I: Integration of cruise and en-route descent (Tests conducted using airlines (B-737, A-300, DC-10) flight simulators)". EUROCONTROL Report 832028, December 1983.
17. Benoît, A. and Swierstra, S. "Ground-air coordinated control of time of arrival. Part II: Integration of final descent, approach and landing (Tests conducted using SABENA DC-10 and NLR EK-28 aircraft flight simulators)". EUROCONTROL Report, in preparation.

WIND MODELLING FOR INCREASED AIRCRAFT OPERATIONAL EFFICIENCY

by
W. Lechner, R. Onken
Deutsche Forschungs- und Versuchsanstalt
für Luft- und Raumfahrt
Flughafen
3300 Braunschweig
Federal Republic of Germany

1. Introduction

Sufficiently accurate knowledge of the wind in magnitude and direction as a function of airspace position and time is an essential part not only for navigation but also for fuel efficiency and efficient planning and control of air traffic.

The matter of concern in this paper, for instance, is the demand for wind data in four dimensional guidance. The calculation of the three dimensional flight path command subject to time-of-arrival constraints (fourth dimension) requires wind prediction along the complete flight path the aircraft is going to track. Since the flight path is to be calculated and updated on line wind information has to be available at any time for the respective three dimensional airspace.

The wind data, currently available through MET service broadcast are not sufficient when used as single source. These data suffer from small measurement rates with respect to location and time. An actual wind situation can extremely deviate from what the MET service announces. Therefore, complementary or even quite different techniques emerge for wind estimation, in particular using modern aircraft sensor and computer equipment.

The Institute of FLIGHT GUIDANCE of the DFVLR has developed such airborne wind estimation techniques. These were implemented in the automatic digital flight control system of the DFVLR's test aircraft HFB 320 and were flight tested.

In order to demonstrate the efficiency of these techniques, automatic 4D approaches were flown starting at a Metering Fix about 50 nm off the airport. Among other criteria the time-of-arrival errors were examined and compared against the allowance of ± 5 seconds.

2. Wind measurement

There are several methods of wind measurement or anemometry which have been applied and are still in use, like

- ground measurement on buildings or towers by rotating anemometers, for instance.
- radio sounding balloon,
- remote measurement, possibly from the ground, by exploiting the Doppler effect created by returning scatter of signals when transmitted in frequencies ranging between that of sound waves and laser frequencies,
- airborne measurement of the vectors for true airspeed and ground speed vectors with determination of the difference.

Obviously, the ground measurement does not work if wind measurement is demanded at higher altitudes, as it is the case for 4D-guidance.

The MET service balloon measurements provide good information at the time of measurement and in the vicinity of the position where the measurement is taken. However, this is insufficient with regard to a larger extent of airspace and time period to be covered.

The remote measurement methods could principally cover the airspace to a large extent, but are not operationally available, last but not least for cost reasons.

The airborne measurement of ground speed and true airspeed appears to be the remaining most suitable and cost-effective method of wind measurement.

The basic relationship between the horizontal wind vector, the ground speed and the true airspeed vector is usually described by the following equations (s. fig. 1):

$$(1) \quad \vec{V}_W = \begin{bmatrix} V_{WE} \\ V_{WN} \end{bmatrix} = |\vec{V}_K| \begin{bmatrix} \sin x \\ \cos x \end{bmatrix} - |\vec{V}_{TAS}| \begin{bmatrix} \sin \psi_a \\ \cos \psi_a \end{bmatrix}$$

$$(2) \quad |\vec{V}_W|^2 = V_{WE}^2 + V_{WN}^2$$

$$(3) \quad \psi_W = \arctg(V_{WE}/V_{WN}) + \pi$$

However, these equations lead to large deviations from the correct values in case of flight conditions with great values of bank angle ϕ and angle of attack α , in particular, if ψ_a is set equal to the azimuth ψ . By use of equation (4)

$$(4) \quad \psi_a = \psi - \alpha \cdot \sin \phi,$$

substantial improvement compared to plain use of the azimuth ψ can be made. This implies sufficiently accurate information about the angle of attack, though.

Fig. 2 demonstrates this effect. Graph 1 depicts the wind speed determined by $\psi_a = \psi$ for a certain flight section containing curved flight portions. Graph 2 shows the corresponding bank angle time history. One can clearly read the undesirable correlation between bank angle and wind speed looking at the bank angle peaks and wind speed offsets during turns. These errors in wind measurement can amount to 7 kts in speed and to about 15 degrees in direction. Graph 3 shows the result for the wind speed measurement when the correction term of equation (4) is used.

3. Wind prediction techniques

The problem associated to wind arises with regard to 4D-guidance because of the need for predictive data of wind speed and wind direction along the planned flight path. How can the wind measurement be used in the process of predicting the wind along the flight path ahead of the current airplane position?

As treated in the last chapter, there are wind data at disposal from onboard measurements and meteorological services. With a data link available, between ground facilities and aircraft all these existing data could be used for the process of wind prediction. This represents the most ideal case. Without the data link, other prediction methods have to be applied, which have to rely predominantly on either onboard measurements or ground derived data.

These techniques, therefore, can be subdivided in

- Ground facility oriented techniques
- Airborne techniques.

In the course of flight trials for a 4D-guidance system in 1983, different wind prediction techniques of these kinds have been investigated by the DFVLR with respect to their applicability to 4D-guidance. The flight trials were performed with the DFVLR flight test aircraft HFB 320 which is equipped with an automatic digital flight control system including a higher level control mode for 4D-guidance. The measurements required are provided by the onboard sensor system, essentially consisting of a digital air data computer, an INS, rate gyros and accelerometers (fig. 3).

3.1 Ground facility oriented techniques

In case of existence of a data link, all data of interest measured onboard aircraft can be transmitted to a meteorological computer center on the ground. All aircraft capable of measuring airspeed and ground speed can be part of the system. For those aircraft which are not capable of measuring ground speed onboard, radar data might possibly substitute for the lacking airborne information. The ground computer will collect all these data and will generate a complete three-dimensional picture of the local wind in a distinct piece of airspace. This technique is often referred to as "windmapping".

The advantage of this method lies in the fact that the majority of onboard measurements which are available accrue along busy air routes, right in the area where we find the highest demand for accurate wind information with regard to 4D-guidance.

In areas with abundance of wind measurements frequently available in time, only mean values have to be determined, possibly by Kalman filtering. In areas where a small number of measurements are available within a limited time period, inter-/extrapolation techniques can be used. The latter will be described briefly in the following by the technique of using the cubic spline function for vertical interpolation over a small number of wind measurements rather evenly distributed over the entire altitude range of concern.

Depending on the number of measurements ($i_{\max} = N+1$), which are used as pivot points for constructing the curve of the wind profile, the considered altitude range is subdivided in $L_{\max} = N$ segments.

The spline technique enables a smooth third order polynomial transition from one segment of the wind profile to the next.

For the vertical profile of wind speed, for instance, the following relationships are valid for the N segments in case of the cubic spline (s. fig. 4):

$$\begin{aligned}
 (5) \quad \Delta h_L &= h_{i+1} - h_i & L=1, \dots, N \\
 \Delta v_{WL} &= v_{Wi+1} - v_{Wi} & L=1, \dots, N \\
 v_{WL} &= a_L + b_L (h-h_i) & L=1, \dots, N \\
 &\quad + c_L (h-h_i)^2 + d_L (h-h_i)^3 \\
 a_L &= v_{Wi} & L=1, \dots, N \\
 b_L &= -c_L \Delta h_L + \Delta v_{WL} / \Delta h_L - (c_{L+1} - c_L) \Delta h_L / 3 & L=1, \dots, N \\
 c_1 &= c_{N+1} = 0 \\
 c_L &= (y_L - \Delta h_L) c_{L+1} / x_L & L=2, \dots, N-1 \quad i = L-1 \\
 c_N &= y_N / x_N \\
 d_L &= (c_{L+1} - c_L) / (3 \Delta h_L) & L=1, \dots, N \\
 x_L &= 2 (\Delta h_{L-1} + \Delta h_L) - \Delta h_{L-1}^2 / x_{L-1} & L=3, \dots, N \\
 x_2 &= 2 (\Delta h_1 + \Delta h_2) \\
 y_L &= z_L - \Delta h_{L-1} \cdot y_{L-1} / x_{L-1} & L=3, \dots, N \\
 y_2 &= z_2 \\
 z_L &= 3 (\Delta v_{WL} / \Delta h_L - \Delta v_{WL-1} / \Delta h_{L-1}) & L=3, \dots, N
 \end{aligned}$$

The evaluation of a great number of wind profiles proved $N = 3$ to be sufficient for the altitude range of FL30 through FL100. Surplus measurements are used for enhancing the measurement quality of the selected altitude pivot levels.

During the flight tests, this technique was simplified in a certain way with respect to the acquisition of wind measurement data, as being utilized as pivot values. These wind data were measured during the climb phase the aircraft went through before the 4D approach test flight was started, and were stored in the onboard computer. Table 1 shows four flight test results of time control achieved when this technique was applied.

3.2 Airborne techniques

Two prediction techniques of airborne type were investigated:

1. Wind prediction through interpolation on the basis of a simplified model.
2. Wind prediction through extrapolation on the basis of the complete wind model the parameters of which are estimated by a Kalman filter.

3.2.1 Wind modelling

The prediction of the wind along the flight path ahead of the current airplane position can in principle, work by inter/extrapolation on the basis of a sufficiently valid description of the mean wind as a function of the airspace location relative to the current position of the airplane, the so-called wind model. By the wind model existing knowledge about the physics of wind can be exploited. The operational use of mean wind data for 4D-guidance calculations does not provide significant shortcomings. High frequency wind variations cannot deteriorate significantly the time control accuracy which is explicitly required at the end point of the 4D-flight phase.

The structure of the wind model can be assumed to be the following:

$$(6) \quad V_{WM} = V_W(h_0) (h/h_0)^D + g_{SE} (x_E - x_{EP}) + g_{SN} (x_N - x_{NP})$$

$$(7) \quad \psi_{WM} = \psi_W(h_0) + g_{RE} (x_E - x_{EP}) + g_{RN} (x_N - x_{NP}) + g_{RH} (h - h_0)$$

$$h \geq h_0$$

Here wind speed in a certain altitude is given by known wind data at a reference altitude h_0 multiplied by an exponential function of altitude. The additional terms in eq.6 are accounting for wind changes within the same altitude by gradients g_{SE} and g_{SN} as a function of the distance of the current airplane position (x_{EP}, x_{NP}) . The model of wind direction is assumed linear in all three dimensions of the geodetic coordinate system with the gradients g_{RE} , g_{RN} and g_{RH} .

This type of exponential model for wind speed and linear model for wind direction as a function of altitude was validated by a large number of wind profiles recorded by the Hanover Airport MET service up to 3000 m altitude (FL 100). The statistics of these recordings are shown in table 2.

3.2.2 Filtering for wind model enhancement

Wind measurements are noisy, partly caused by measurement errors. Since it is difficult to tell apart measurement errors from high frequency wind changes, only mean wind data as a function of airspace position and time will be used for the 4D-guidance calculations. As mentioned before, this is sufficient

to comply with the 4D-guidance time control requirements, if these mean data can be made available and if they can be predicted at locations ahead of the current airplane position. This necessarily leads to filtering of the measured wind data. The wind model serves as part of the filtering to structure the mean wind relationships. The parameters of this model are the crucial quantities to be determined as well as the errors of these model parameters. The best suited filter for this purpose is the Kalman filter which processes the difference between the low frequency wind model output and the high frequency wind measurement signal (fig. 5). The Kalman filter provides an estimate of the errors of the wind model the behavior of which is described by linear error equations, the so-called error model, derived by partial differentiation of the wind model equations. This error model is part of the Kalman filter forming the "system matrix" Φ .

The error state vector with regard to wind speed has got 5 components: error of wind speed, error of reference wind speed, error of the exponent p and errors of horizontal gradients g_{SE} and g_{SN} . The error state variables with regard to wind direction are the error of wind direction itself and the error gradients g_{RE} , g_{RN} and g_{RH} . The system matrices Φ_S and Φ_R for wind speed and wind direction respectively are the following:

$$(8) \quad \Phi_S = \begin{bmatrix} 1 & \Phi_S(1,2) & \Phi_S(1,3) & \Phi_S(1,4) & \Phi_S(1,5) \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\Phi_S(1,2) = \Delta x_E(k)$$

$$\Phi_S(1,3) = \Delta x_N(k)$$

$$\Phi_S(1,4) = p \cdot [h(k)/h_0]^p \cdot [\Delta h(k)/h(k)]$$

$$\Phi_S(1,5) = v_W(h_0) \cdot p \cdot \ln[h(k)/h_0] \cdot [h(k)/h_0]^p \cdot [\Delta h(k)/h(k)]$$

$$(9) \quad \Phi_R = \begin{bmatrix} 1 & \Phi_R(1,2) & \Phi_R(1,3) & \Phi_R(1,4) \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\Phi_R(1,2) = \Delta x_E(k)$$

$$\Phi_R(1,3) = \Delta x_N(k)$$

$$\Phi_R(1,4) = \Delta h(k)$$

In order to estimate the wind model errors with sufficient accuracy, additional information for the filtering is needed about those errors which are not described by the error equations. These unmodelled errors are accounted for by the system noise matrix. Also the statistics of the measurement errors, like INS and air data sensor errors, have to be taken into account.

This information is derived from data like those in table 2 and known statistical error models of the sensors. The measurement noise was defined by a discrete function of the bank angle in order to take into account the increase in error magnitude during turns.

The Kalman filter can then be used in two ways, first as smoothing filter and second as estimation filter of the wind model parameters with regard to wind prediction along the 4D-flight path ahead. Both applications, in particular the latter, will be shown in the following chapter.

3.2.3 Wind prediction through interpolation

This wind prediction technique yields the vertical wind profile assuming no wind changes without change of altitude. This is in conformity with the wind model described before by setting the horizontal wind gradients equal to zero. Two measurements ahead at different altitudes, serve as input data to determine the unknown wind model parameters. For this purpose, also MET service data can be utilized in particular for the extreme (low) altitude at which the aircraft will eventually arrive from its current position. The local wind measurement is taken by the onboard sensor system.

For a typical 4D approach the altitude changes from FL100 down to FL30 at the Merge Gate, the point above the extended runway centerline the 4D-flight path is destined for. As the aircraft is approaching the Merge Gate the interpolation interval decreases and the wind model is becoming more and more realistic.

Fig. 6 depicts the wind profiles derived by this technique at four different stages of flight progress. Accuracy demands are prevailing in particular after the point where the last corrective change of the flight path is possible, not far from the Merge Gate.

Table 3 contains the 4D-flight test results for the time-of-arrival error as a result of this technique of wind prediction.

Fig. 7 shows the time history of the TOA error for flight number 14, calculated on-line on the basis of the navigation and wind measurement data available at that time instant. The poor modelling of the actual wind profile during the first section of the 4D approach causes the increase in time error as can be seen in the plot. If the time error exceeds the error limits, as shown in the figure (the limits are decreasing with time from ± 20 sec down to ± 5 sec), the 4D-guidance mode compulsorily generates a corrected flight path in order to compensate for the time error. However, during the last section of the 4D-approach time error compensation by flight path modifications is no longer possible. It can be seen from the figure that right at this crucial flight phase the wind modelling is sufficiently accurate. This leads to the conclusion, that the feasibility of this technique depends very much on the accuracy of the wind data at the Merge Gate, the final point of the 4D-landing approach flight path.

3.2.4 Wind prediction through extrapolation

The principle of this technique is depicted in fig. 8. The wind model calculation is enhanced by a Kalman filter which produces an estimate of the wind model parameter errors. Besides the wind model output signals and actual onboard wind measurements information about the noise levels of the wind model calculation and the onboard wind measurements is to be provided for the Kalman filtering.

Fig. 9 presents an example for this wind prediction technique. It shows the measured values as well as the predicted ones for wind speed and wind direction. The time at the beginning of this plot is the actual point of time of the flight test. The prediction from this time on is illustrated in this figure (solid line) to demonstrate the prediction performance. These predicted values are to be compared with the corresponding values measured later on (dots) when the aircraft follows the commanded 4D-flight path which was calculated on the basis of wind prediction. Obviously, the accuracy of prediction declines with increasing prediction time but the errors are still below 7 kts and 10 degrees for wind speed and wind direction respectively after 16 minutes of subsequent flying time. The particular advantage of this technique, of course, is the capability of self-dependent onboard wind prediction.

Because this technique was of major interest, a total of 8 flight tests were conducted. The time-of-arrival errors of these flights are listed in table 4. Again, the results comply well with the requirements.

4. Conclusion

In the course of the development of an automatic four dimensional guidance system for efficient control of operational air traffic approaching the airport a function for, prediction of wind speed and wind direction along the calculated flight path had to be performed with sufficient accuracy.

Wind prediction techniques were investigated for this purpose. It was made use of the wind measurements available on the ground and/or onboard the aircraft as well as of a model describing the local wind situation around the actual aircraft position where actual wind information is at hand. In particular, the wind model could be utilized for extrapolation purposes with regard to prediction and could be effectively implemented in conjunction with a Kalman filter for wind model error estimation.

All techniques described, differing in their technical and operational features, complied with the accuracy requirements, at least under the conditions of the flight test experiments conducted at the DFVLR with the test aircraft HFB 320.

5. References

- [1] Gelb, A.; Applied Optimal Estimation, MIT Press, 1974.
- [2] Hemesath, N.B., Bruckner, M.H., Krippner, R.A.; Three and Four Dimensional Area Navigation Study, FAA-Report, FAA-RD-74-150, June 1974.
- [3] Pecsvaradi, T.; Four-Dimensional Guidance Algorithms for Aircraft in an Air Traffic Control Environment; Ames Research Center, NASA, TN D-7874, Aug. 1975.
- [4] Adam, V., Lechner, W.; Investigations on Four-Dimensional Guidance in the TMA, AGARD Conference Proceedings No. 340, Lisbon, Sept. 82.
- [5] Bisiaux, M., Cox, M.E., Forrester, D.A., Storey, J.T.; Possible Improvements in Meteorology for Aircraft Navigation EUROCONTROL, Doc. 82.20.37, Bruxelles, Nov. 1982.
- [6] Adam, V., Lechner, W.; Guidance and Control Research Flight Testing with HFB 320 Test Aircraft; AGARDograph No. 262, Ground and Flight Testing for Aircraft Guidance and Control Systems, to be published 1984.

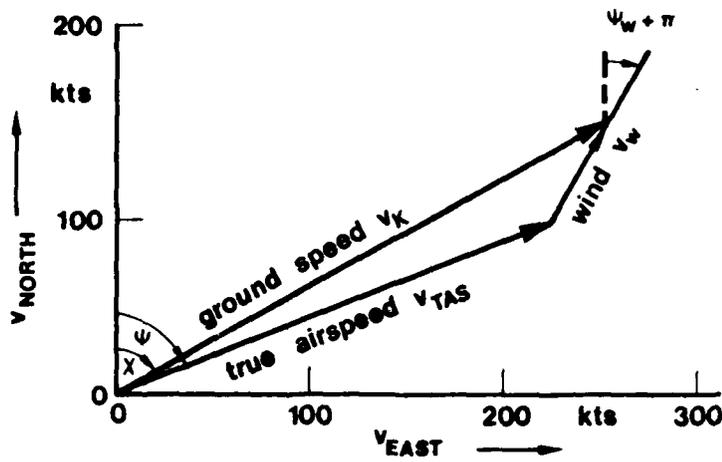


Fig. 1: Basic relationship between true airspeed, ground speed and wind

Graph 1: Measured wind speed (kts) —○—
 Graph 2: Bank angle (deg) —x—
 Graph 3: Corrected wind speed (kts) —+—

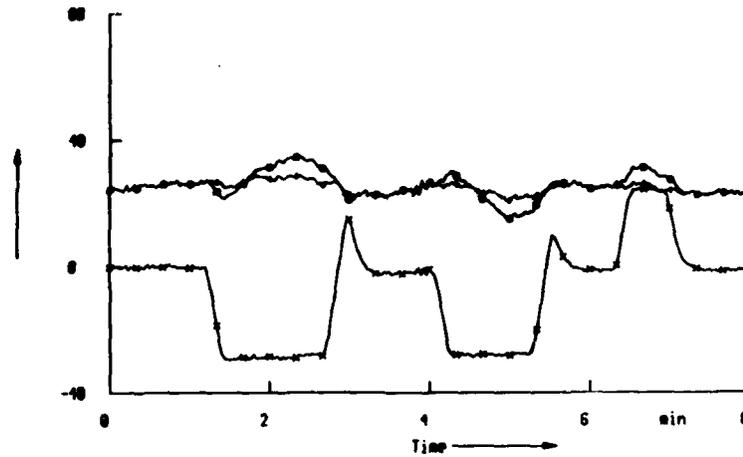


Fig. 2: Improved wind measurements

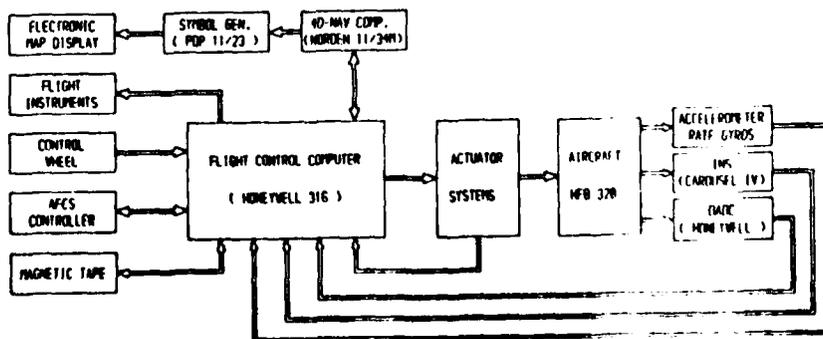


Fig. 3: Structure of the automatic digital flight control system of the HFB 320 test aircraft

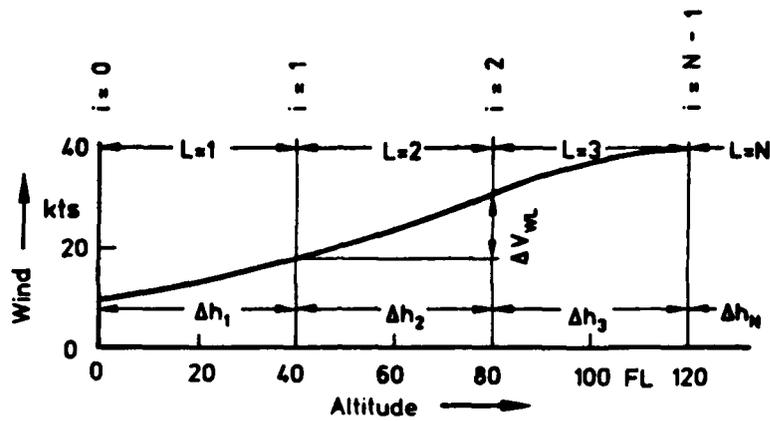


Fig. 4: Definition of a cubic spline function

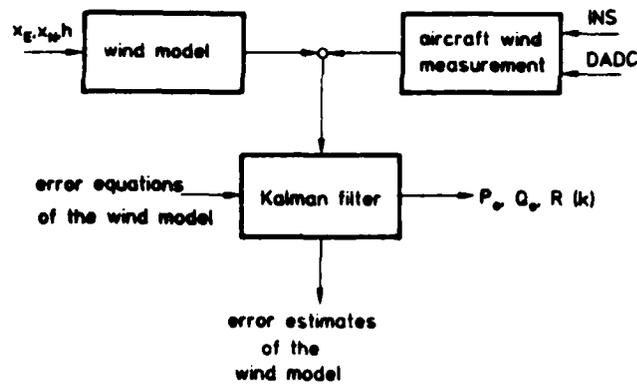


Fig. 5: Principle of Kalman Filtering for wind measurement

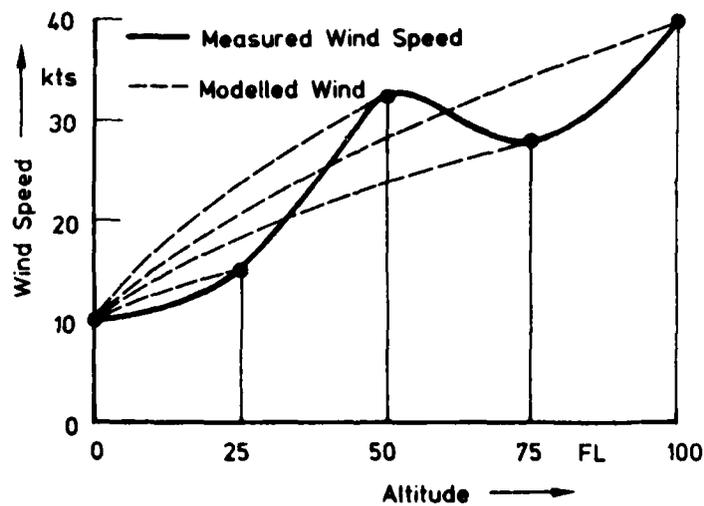


Fig. 6: Windmodelling by interpolation

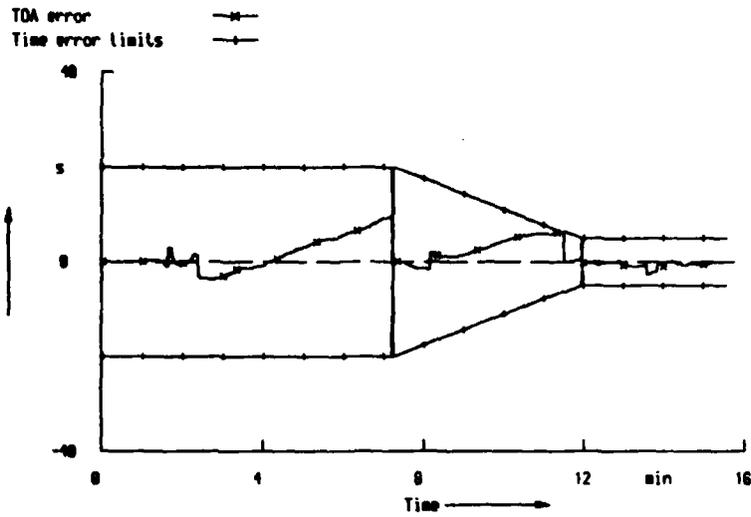


Fig. 7: Time history of a time-of-arrival error

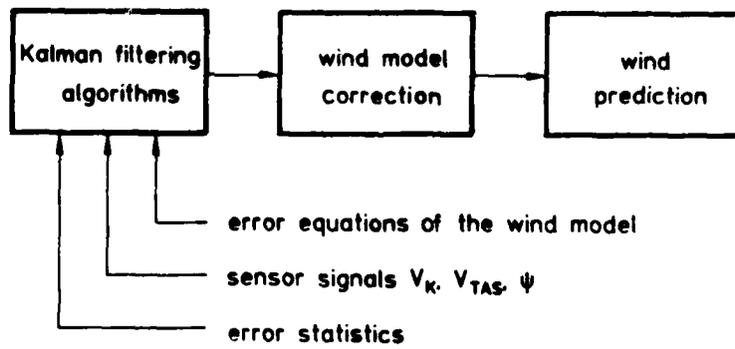


Fig. 8: Principle of wind prediction through extrapolation

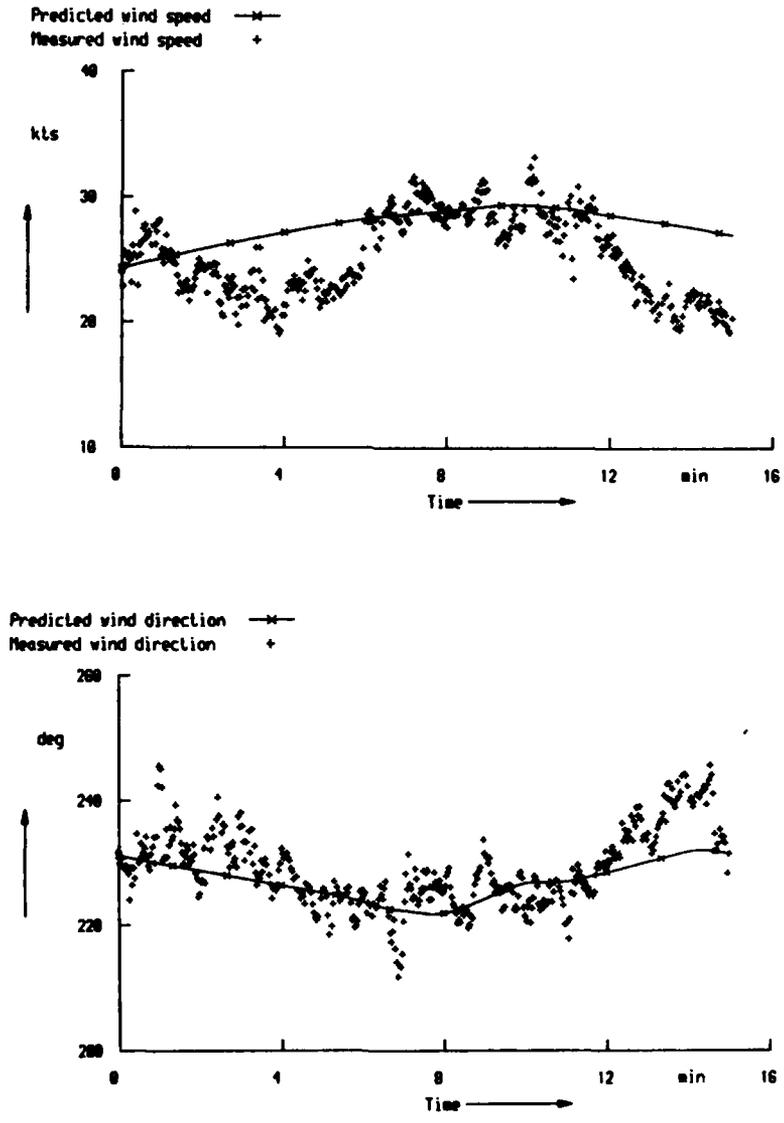


Fig. 9: Example for the wind prediction through extrapolation

| | | | | |
|----------------------|-------------|-------------|-------------|-------------|
| flight number | 3 | 5 | 10 | 13 |
| TOA error | -2.5 | -0.6 | -0.5 | -1.7 |

Tab. 1: TOA errors in seconds (cubic spline function)

| | mean value | standard deviation |
|-----------------------------|--------------------|---------------------------|
| wind speed error | 2.2 kts | 5.9 kts |
| wind direction error | 6.6 deg | 16.0 deg |
| exponent p | 0.345 | 0.24 |
| vertical gradient | 0.27 deg/FL | 0.7 deg/FL |

Tab. 2: Statistical analysis of wind profiles

| | | | | |
|----------------------|------------|------------|-------------|------------|
| flight number | 4 | 8 | 11 | 14 |
| TOA error | 0.7 | 4.0 | -1.4 | 0.1 |

Tab. 3: TOA errors in seconds (wind prediction through interpolation)

| | | | | | | | | |
|----------------------|------------|------------|-------------|------------|------------|-------------|-------------|------------|
| flight number | 1 | 2 | 6 | 7 | 9 | 12 | 15 | 16 |
| TOA error | 4.4 | 3.1 | -2.3 | 2.9 | 0.2 | -3.7 | -1.0 | 2.0 |

Tab. 4: TOA error in seconds
(wind prediction through extrapolation)

A METHOD OF ESTIMATING AIRCRAFT ATTITUDE
FROM FLY BY WIRE FLIGHT CONTROL SYSTEM DATA

by

Reginald J.V. Snell
British Aerospace PLC
Aircraft Group, Warton Division
Warton Aerodrome, PRESTON PR4 1AX
England

SUMMARY

The trend towards the use of Active Control Technology (ACT) to control the aircraft flight path has led to an increasing investment in flight path sensors and associated computing. In an attempt to offset the cost, and at the same time save weight, a study was conducted at British Aerospace, Military Aircraft Division at Warton into using this data to extract aircraft attitude. If this could be achieved it would not only avoid the requirement for supplementary hardware in the form of an Attitude References System and Artificial Horizon etc., but could endow the source of attitude with the high integrity and availability associated with flight control systems. This would be appreciated by pilots who were pressing for automatic cross referencing of flight reference data to minimise cockpit work load in flight situations where the pilot needs to fly 'eyes out' of the cockpit as much as possible.

Following initial studies with flight recorded data, an experimental application was flown on the ACT 'Fly-by-Wire' Jaguar Demonstration aircraft using digital data originating in the Flight Control System. The flight results have demonstrated good accuracy and stability in the general course of flying. During violent manoeuvres, errors in pitch peak at 2° to 3° but are generally insignificant by completion of the manoeuvre.

If attitude can be established with sufficient accuracy, it would be feasible to extract aircraft magnetic heading from a three axis magnetometer. Heading, together with attitude and air data information would provide the basis for a dead reckoning navigation system. It is believed that such a basic system could be aided by a continuous terrain profile matching system to provide high accuracy navigation. Consequently, the method not only offers a source of high integrity attitude but could also serve as the basis for a low cost, highly accurate navigation system suitable for operation over digitally mapped terrain, either as a stand alone system or a back up for cross monitoring an inertial navigation system etc.

BACKGROUND

The motivation for the development of a technique to extract attitude from the Flight Control System arose out of a study for a low cost, lightweight combat aircraft. To increase tolerance to higher normal acceleration levels in combat, the pilot's seat was to be more reclined than conventionally which had the effect of reducing the visual area available for cockpit instrument display. Because of the limited display space, there was pressure to reduce the number of secondary instruments used by the pilot for cross checking. This was in line with a demand from pilots for automatic cross referencing, to reduce pilot work load in flight situations where they need to fly eyes out of the cockpit. The standby displays are chosen for their high availability derived from design simplicity and maturity. Consequently any single source of data must have both integrity (to avoid cross monitoring) and high availability (to replace standby instruments).

The aircraft was to have a digital flight control system with multiple lanes of sensors to establish both high integrity and availability. Therefore any function which used flight control system data might be endowed with integrity and availability. As the flight control system uses gyroscopes and attitude systems also use gyroscopes, could attitude be a suitable subject for investigation?

DERIVATION OF ATTITUDE

The traditional attitude reference system in its most basic form comprises a gyrostabilised reference axis aligned to the vertical and corrected for drift by using accelerometers acting in the sense of a plumb bob. The accelerometers of course cannot distinguish between gravitational and inertial acceleration and consequently their vertical reference is only dependable over periods of time of sufficient duration wherein the influence of manoeuvre acceleration is small in relation to the sustained gravitational acceleration. Accordingly, the coupling between the accelerometers' and gyroscopes' vertical references must be comparatively weak to avoid significant attitude errors during periods of severe aircraft manoeuvres. For this reason, the random drift rate of the attitude reference gyroscopes is traditionally not greater than several degrees per hour.

The gyroscopes in current ACT systems have random drift rates in the order of several degrees per minute or nearly two orders greater than the requirement for traditional attitude reference. Consequently, in order to employ ACT gyroscopes for the derivation of attitude, there must be a closer coupling with the long term vertical reference necessitating a better vertical reference than from the accelerometers alone.

As the accelerometers cannot distinguish between gravitational and inertial accelerations, then the vertical reference could be improved by correcting the accelerometer measurements for inertial acceleration.

The inertial acceleration acting on the airframe is of course produced by change of magnitude and direction of aircraft velocity. This acceleration can be expressed in terms of its components acting along the aircraft reference axes.

Figure 1 shows the traditional right handed set of axes x, y and z, centred on the aircraft centre of gravity, with x lying along the aircraft fuselage. The components of aircraft velocity V along these axes are respectively V_x , V_y and V_z and the rotation rates about the axes respectively P, Q and R. The inertial acceleration along an axis is then the sum of an acceleration caused by change of magnitude of velocity along the axis and two centripetal components induced by velocities and rotations about the two adjacent axes as shown in Figure 2 for the x axis. The inertial accelerations for the three axes can then be expressed as

$$\begin{aligned} \ddot{x}_i &= \dot{V}_x + V_z Q - V_y R \\ \ddot{y}_i &= \dot{V}_y + V_x R - V_z P \\ \ddot{z}_i &= \dot{V}_z + V_y P - V_x Q \end{aligned}$$

Looking to the Flight Control System (FCS) for sources of data to evaluate these expressions, then clearly the rotational terms P, Q and R would be available from the FCS gyroscopes. Other sensor data used by the FCS include the air data measurements, pitot static pressures, and angles of attack α and side slip β . In conjunction with outside air temperature measurements, True Air Speed can be evaluated from the pressure measurements to represent aircraft speed V and the components of this velocity along each axis deduced from the α and β measurements. As pitot probes generally have a high pressure recovery over a wide angle of attack, then it may be taken that True Air Speed represents the velocity V of the aircraft in the air mass. Referring to Figure 3, the resolution of V onto the x axis will then correspond to the direction cosine

$$l = (1 + \tan^2 \alpha + \tan^2 \beta)^{-1/2}$$

so that

$$\begin{aligned} V_x &= V l \\ V_y &= V l \tan \beta \\ V_z &= V l \tan \alpha \end{aligned}$$

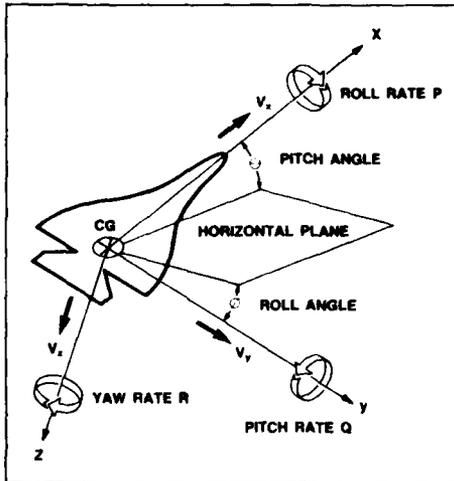


Figure 1 Aircraft Reference Axes and Velocity Notation

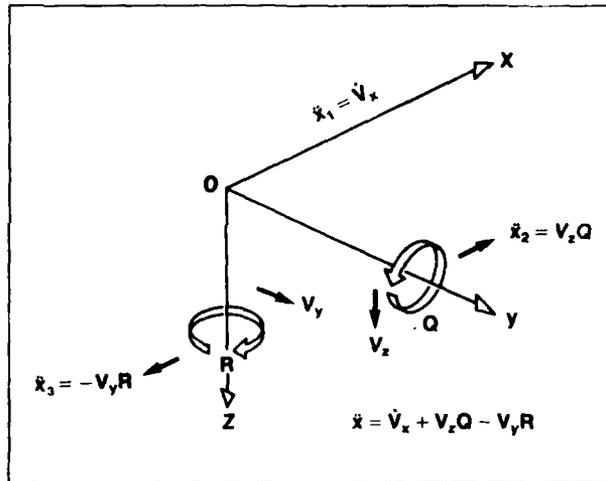


Figure 2 Accelerations along the X Axis

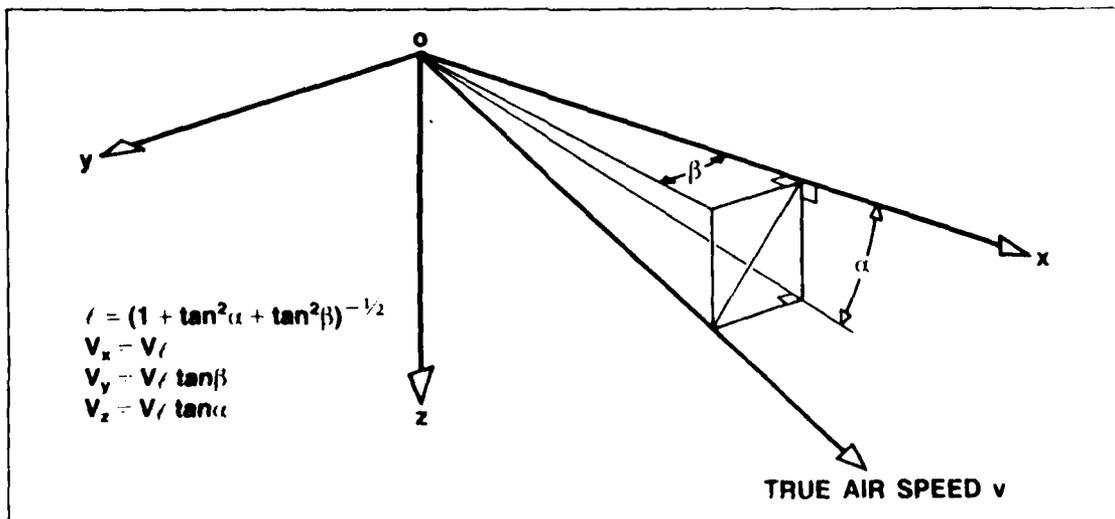


Figure 3 Resolution of True Air Speed onto X Y Z Axes

This method of estimating inertial acceleration is based upon the use of air speed and the implications require examination. Now, fundamentally there is no measure of absolute velocity. All velocities must be relative to some arbitrary reference frame, the most convenient usually being the earth's surface. In formulating accelerations, due account must be made for any acceleration of the reference frame itself. For instance, rotation of the Earth will induce Coriolis terms which must be allowed for in Inertial Navigation systems.

Therefore in using the air mass as a frame of reference, the wind speed is not directly of concern but change of wind speed is pertinent. In general, significant air mass acceleration appears as gusts of comparatively short duration and the resulting errors may be amenable to smoothing.

The next question is, what is the potential accuracy of such a system? In general, attitude accuracy is more important in sustained steady manoeuvres, level flight, descents and turns to avert excessive deviation of the flight path. As rapid manoeuvres are relatively short term then larger errors may be tolerated provided there is no residual error at the end of the manoeuvre. Consider the steady state case where in principle there would be no differentiation of the components of air speed along each axis. The accuracy would depend upon the errors in accelerometer and gyroscope measurements.

The ACT Jaguar Digital Flight Control System uses 12 bit digitisation of sensor data and it is assumed that projected systems will also use at least 12 bits if not extended to 14 bits. If the accelerometers on the longitudinal and transverse axes have a range of ±2g then the least significant bit (LSB) with 12 bit digitisation, would be 0.001g, amounting to an attitude error of 0.06 degrees.

The corresponding errors attributable to the gyroscopes will arise from centripetal accelerations induced by velocities along adjacent axes.

Assuming negligible sideslip in the steady state case, the centripetal acceleration error along the longitudinal or x axis would be due to velocity along the normal or z axis and pitch rotation error. Typical pitch gyroscope range would be ±60°/sec. or a LSB of 0.03°/sec. for 12 bit digitisation. At Mach 0.9 or 300 metres/sec. and, say, an angle of attack of 2 degrees, the velocity along the z axis would be 10.5 m/s and the acceleration induced along the x axis would be 0.0055 m/sec.², amounting to a pitch error of 0.03 degrees.

The centripetal acceleration error along the transverse or y axis will be due to velocity along the x axis and yaw rotation error. Again, the yaw gyroscope operating range will be about ±60°/sec. and LSB 0.03°/sec. Therefore at Mach 0.9 with a velocity along the x axis of 300 m/s, the acceleration error will be 0.157 m/sec.² amounting to a bank error of 0.92 degrees.

Therefore the intrinsic pitch error due to digitisation will be of the order of 0.1 degree and the bank error 1 degree. Pitch is always the more critical channel and larger errors are acceptable in bank.

There is a further acceleration term to consider, induced by travelling over the curved surface of the earth. This is the Coriolis term acting at right angle to the direction of travel, expressed as:

$$- 2V\Omega \sin \lambda \quad - \frac{V^2}{R} \tan \lambda \cdot \sin \psi$$

where V is ground speed, ψ is track angle, Ω Earth's rotation rate, R Earth's radius and λ latitude angle. The term can be computed from navigation data, but if not allowed for, it would generate errors in bank of the order of 0.2° to 0.3° at Mach 0.9 at Northern European latitudes.

Therefore, having realised that the magnitude of windspeed was not a critical factor and that the system had the potential for adequate intrinsic accuracy, the concept was tested using flight recorded data. Such data was already available from pre-production Tornado aircraft, instrumented for flight handling clearance. The instrumentation included a side slip and three axes of accelerometers not normally fitted to production aircraft. Further, the records of angles of attack and side slip and the accelerometer reading were corrected to the aircraft's centre of gravity location. The instrumentation records also included pitch and bank measured by the aircraft's inertial navigation system, providing a convenient datum for assessing the accuracy of the technique. The data is recorded digitally but only to 10 bits, which would coarsen the attitude estimates in relation to their potential use with a flight control system.

Initial explorations with this data were aimed at calculating attitude directly from the components of gravitational acceleration on each axis \ddot{x}_k , \ddot{y}_k and \ddot{z}_k calculated from the difference between the measured total acceleration and estimated inertial acceleration on each axis.

$$\text{pitch angle } \theta = \text{arc sin } \left(\frac{\ddot{x}_k}{g_1} \right)$$

$$\text{bank angle } \phi = \text{arc tan } \left(\frac{\ddot{y}_k}{\ddot{z}_k} \right)$$

Where g_1 is the local gravitation acceleration taken as

$$g_1 = \sqrt{\frac{\ddot{x}_k^2}{R} + \frac{\ddot{y}_k^2}{R} + \frac{\ddot{z}_k^2}{R}}$$

to avoid numerical problems when \ddot{x}_k/g_1 approaches unity.

This proved to be excessively noisy due, as one might expect, to differentiation of the component accelerations measured on each axis - a large part of the noise being due to digital quantisation.

Several types of filter were investigated to reduce the noise induced by differentiation. The most suitable was found to be a recursive differentiator and filter of the type:

$$\dot{x}_i = A(x_i - x_{i-1}) - B\dot{x}_{i-1} - C\dot{x}_{i-2}$$

where
$$A = \frac{1 + B + C}{\Delta T}$$

The values of B and C were selected, in relation to the sampling interval ΔT to give a cut off a little over 1 Hz, this being the order of pitch response frequency of typical fighter aircraft. The filter produces a substantially linear phase lag with frequency up to about 1 Hz so that all frequencies less than 1 Hz have a similar time lag.

To synchronise the rotational acceleration terms, they were deliberately delayed by the amount of the filter time delay before being added to the linear acceleration terms produced by differentiation. The resulting attitude estimates were now stale, and were brought up to current time by adding an incremental change of attitude deduced by integrating gyro body rate measurements over the filter time delay interval.

The resulting prediction of attitude agreed quite well long term with the inertial navigation reference but was still far too noisy to be used alone.

The other source of attitude is that produced by integration of the gyroscope body rates. This estimate is smooth but prone to long term drift. The magnitude of the drift was difficult to assess. Comparing this source of attitude with the inertial reference, indicated a divergency of the order of the LSB in the 10 bit instrumentation records, implying that the intrinsic drift of the gyroscopes was much less.

The situation was obviously a suitable candidate for a Kalman Filter wherein the long term stability of the noisy attitude estimates can be combined with the smooth but drifting estimates derived from integration of gyro body rates. A continuous-discrete extended Kalman Filter with pitch and roll as the state variables was duly implemented. The attitude estimates now proved to be sufficiently smooth and accurate to warrant further work on the concept, culminating in a flight demonstration. Because the method was based upon building up the gravitational acceleration components from the components of inertial acceleration, it became termed the Synthetic Attitude Technique.

FLIGHT DEMONSTRATION

The obvious vehicle for a flight demonstration was the ACT Jaguar demonstration aircraft flying at Warton with a digital flight control system. The sensor data available within the FCS included pitot-static pressures, angles of attack and side slip together with all three axes of gyroscope measurement of aircraft body rotation rates. The only data lacking was the acceleration measurements along the aircraft axes. A further important consideration was that the aircraft FCS ground rig was sited at Warton and could be used to develop and clear the technique before embarking upon flight trials.

The most appropriate place to invest the software for computing attitude would be in the FCS computers to thereby acquire integrity and availability. To pursue this policy on the demonstration aircraft would cause inevitable delays to the main flight trials programme and was therefore untenable. Similarly any attempt to display or involve this source of attitude with other avionic subsystems was discarded to avoid delays and the expense incurred in clearing flight safety implications. The outcome was that a dedicated processor would be required which simply fed data into the Flight Data Recording System. Access to data would be a straightforward matter of tapping into the 25 Hz digital data stream from the FCS computers to the Flight Data Recording System.

A processor was constructed by the Electronics Department at Warton, and designed to receive 16 bit data from the FCS computers, convert analogue data from the accelerometers and a total temperature probe to 12 bits, perform computations in 16 bits and output data in analogue form for flight recording. In addition to the fundamental attitude computation, the processor had to derive true air speed from the pitot static and total temperature measurement and correct accelerometer measurements to the aircraft centre of gravity position. Limitations in computing capacity prevented inclusion of corrections for compressibility effects on the pitot static measurements in the transonic region. The offset of the accelerometers from the aircraft centre of gravity (CG) were based on a mean CG position in flight.

Acceleration measurements were provided by a set of three Systron Donner accelerometers type 4311 with a performance compliant with 12 bit digitisation. The range on the x and y axes was +2g and -2g to 10g on the z axis. The accelerometers were mounted on a common bracket to maintain orthogonality and mounted about one metre from the aircraft's centre of gravity, which was as close as the installation would allow. The mounting site was in the spine which facilitated alignment of the axes of these accelerometers to the aircraft axes.

FLIGHT RESULTS

In assessing the performance of the technique, the essential feature should be small error in sustained flight conditions to avoid pronounced deviation of the flight path from the intended direction. Small error excursions could be tolerated provided they were of short duration. Larger errors can be tolerated in highly dynamic manoeuvres because of their shorter duration provided there was a return to normal conditions within a very short time of completing the manoeuvre.

The degree of success of the technique with respect to these criteria may be gauged from the specimen time histories shown in Figures 4 to 11. These histories show the difference between synthetic attitude estimated in flight and the corresponding Inertial Platform attitude, subtracted post flight from 10 bit digital flight records. The flight began with low dynamic flight manoeuvres and progress through pitch manoeuvres, high g turns and rapid rolls to represent increasingly more severe dynamic conditions.

In the low dynamic conditions, a 20° banked turn and a landing approach descent Figures 4 and 5, the mean pitch error is zero. Peak pitch error excursions are only $\frac{1}{2}^\circ$ in the turn and there is one instance of a peak of 1.2° in the descent. The bank error in the turn shows a bias of about -2° with a peak error of -4° . In the descent the mean bank error is zero but the peak error is 3° where 2° is exceeded for about one second.

Figures 6 and 7 show some pitch manoeuvres with wings level pulling some 2 and 2.6g. In both cases the pitch error is contained within 1° . In Figure 6 the bank errors also lie within 1° but in Figure 7 the errors peak at $\pm 6^\circ$ at the point where the pitch rate is greatest.

Two level turns at 3g and 6g respectively are shown in Figures 8 and 9. In both cases the pitch errors are within $\pm 1.5^\circ$. The bank errors peak at $+6^\circ$ and -8° in the 3g turn and $+5^\circ$ and -4° in the 6g turn at maximum bank rate. In the sustained part of the 6g turn the bank error is generally less than 3° .

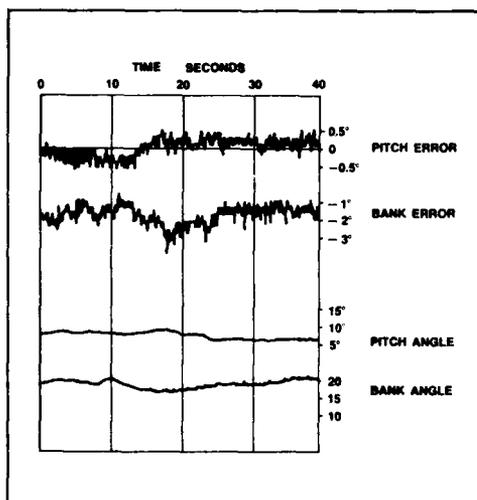


Figure 4 2g Turn

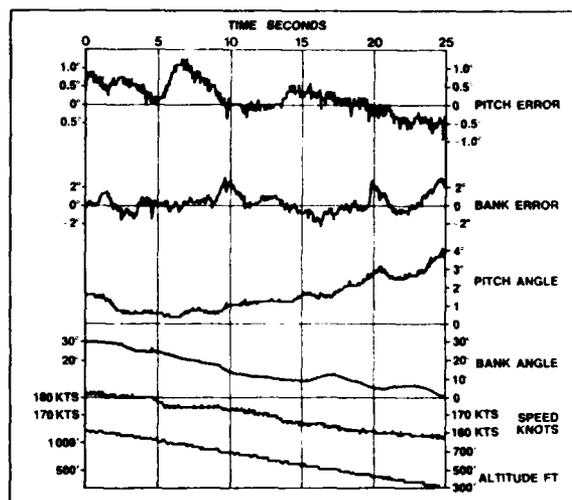


Figure 5 Landing Approach

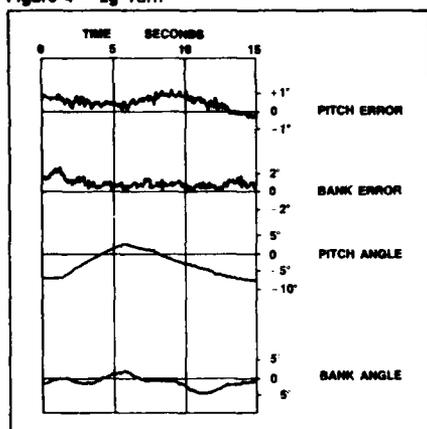


Figure 6 Straight 2g Pull

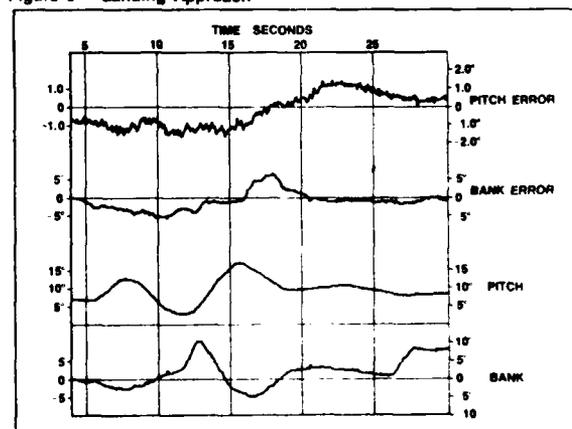


Figure 7 Rapid Pitch Manoeuvres

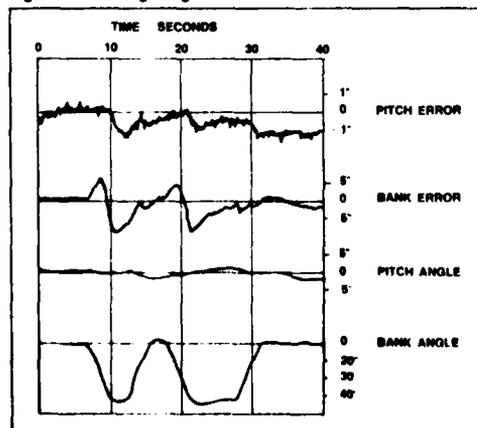


Figure 8 3g Turns

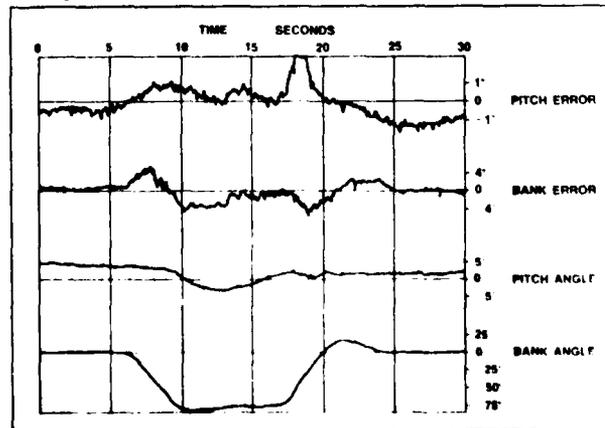


Figure 9 6g Turn

The final two manoeuvres are with high roll rates, a 2g Wind-up Turn (WUT) and some rapid rolls, Figures 10 and 11. The pitch error in the WUT is substantially zero with a peak error of 0.8°. However, although the average bank error approaches zero, there is a very large error of 20° at the maximum roll rate. The rapid rolls exhibit similar characteristics. The pitch errors are within ±1.5° but the bank errors attain 25°, again at the maximum rate of change of bank angle.

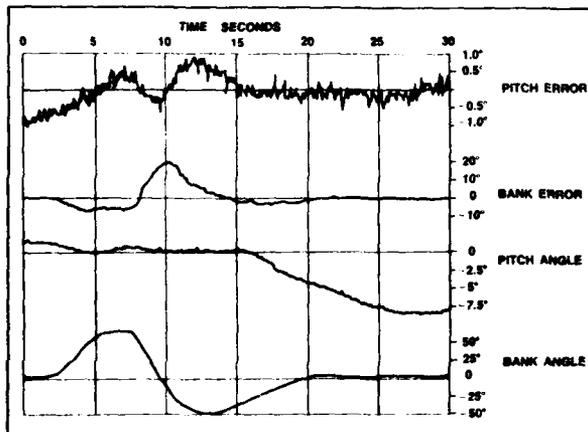


Figure 10 2g Wind-Up Turn

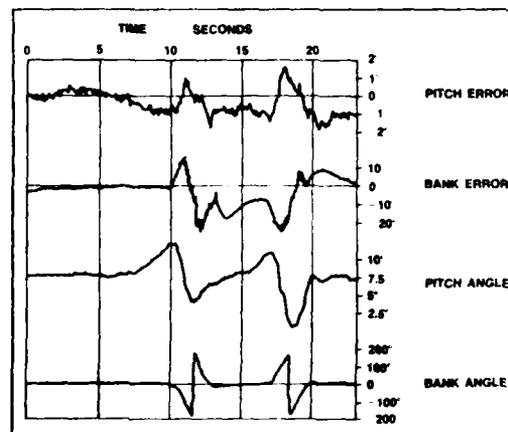


Figure 11 Rapid Rolling

To sum up then, the mean pitch error is zero especially in gentle manoeuvres with some maximum errors generally not exceed 1 to 1½ degrees. The bank errors are larger, but still small in gentle manoeuvres. In high dynamic manoeuvres, the bank errors are pronounced at maximum pitch and bank rates but return to negligible generally before completion of the manoeuvre.

Thus at the present state of development, the flight results may be claimed to be most encouraging. The latest analysis of flight data has shown that the large bank errors could be reduced by changing the coefficients in the velocity differentiator routine and the associated time delay corrections.

It is also believed that some further reduction in error could be achieved by placing more weighing on the gyroscope attitude reference within the Kalman Filter. However the random drift rate of the gyroscopes, especially as they are a consolidated mean of four gyroscopes on each axis, is much less than the least

significant bit of the flight recorded data and therefore this line of investigation is impractical with the current standard of flight recording.

MAGNETIC HEADING

Although the Synthetic Attitude Technique is primarily concerned with attitude, it can be made to fulfil a greater role in conjunction with other technical developments.

One such development is the three axis magnetometer or three axis flux detector (3AFD). If the attitude of each sensing axis of the 3AFD is known, the horizontal component of the Earth's magnetic field, that is magnetic north, can be resolved out of the magnetometer readings to give aircraft magnetic heading. As with the Synthetic Attitude Technique, the FCS gyroscopes can be used to smooth out noise. Heading rotation rate can be resolved from the FCS gyroscopes and Kalman Filtered with basic magnetic heading from the magnetometer.

The attraction of the technique is that it is not prone to errors in turns as is the earlier pendulously mounted two axis magnetometer as it swings out of the horizontal plane into the bank plane.

Exploratory flight trials are being undertaken on a three axis magnetometer, also on the ACT demonstrator Jaguar. An experimental 3AFD constructed by BAE Dynamics Group, Bracknell Division, is mounted near the top of the fin in place of the forward passive radar warning detector head with its three sensing axes aligned to the aircraft axes. Initial calibration of the device has been made with reference to Inertial Navigator heading and attitude to resolve published Earth's magnetic field data onto each axis of the magnetometer for comparison with measured field strength.

The true component of the Earth's field on each axis can be expressed in terms of the total measurements x_m , y_m and z_m to account for axis misalignment, hard iron single cycle and soft iron two cycle effects, in the form:

$$H_x = a_1 x_m + b_1 y_m + c_1 z_m + d_1$$

$$H_y = a_2 x_m + b_2 y_m + c_2 z_m + d_2$$

$$H_z = a_3 x_m + b_3 y_m + c_3 z_m + d_3$$

The corrected measurements H_x , H_y and H_z are then resolved onto the horizontal plane, along and across aircraft heading, using pitch and bank attitude as follows.

$$H_A = H_z \cos \theta + \sin \theta [H_y \sin \phi + H_x \cos \phi]$$

$$H_x = H_y \cos \phi - H_z \sin \phi$$

The horizontal component of the Earth's magnetic field is therefore:

$$H = \sqrt{H_A^2 + H_x^2}$$

and aircraft magnetic heading becomes

$$\gamma_H = \arccos\left(\frac{H_A}{H}\right) \text{ or } \arcsin\left(\frac{-H_x}{H}\right)$$

The traditional ground compass swing will exercise the x and y axes of the magnetometer from the positive to the negative magnitude of the Earth's horizontal magnetic field whilst the magnetic field on the z axis remains substantially unchanged. Consequently the device can be calibrated fully only in flight in manoeuvres including steep banks and inverted flight. At the time of writing only part of the flight envelope has been analysed. However, the limited data gathered so far can serve to illustrate the potential of the device.

Figure 12 shows a post flight estimate of magnetic heading in relation to true heading based on flight records from the magnetometer and inertial navigation system heading and attitude. The first manoeuvre is a 60/65° bank turn through 360° of heading followed by a barrel roll and a slow roll. The turn illustrates how, throughout the full 360° heading change, magnetic heading maintains within 1/2° or so of a constant difference with true heading amounting to the local magnetic deviation. In the course of the two rolls there is a deviation from mean magnetic heading of about 4° or so corresponding to rudder application to counter side slip as the aircraft rolls through 90° of bank. The rudder application causes the fin to bend and twist thereby slightly misaligning the magnetometer axes to give rise to the error. The effect is pronounced because the high angle of dip results in a heading error of some 2.5 times the attitude error.

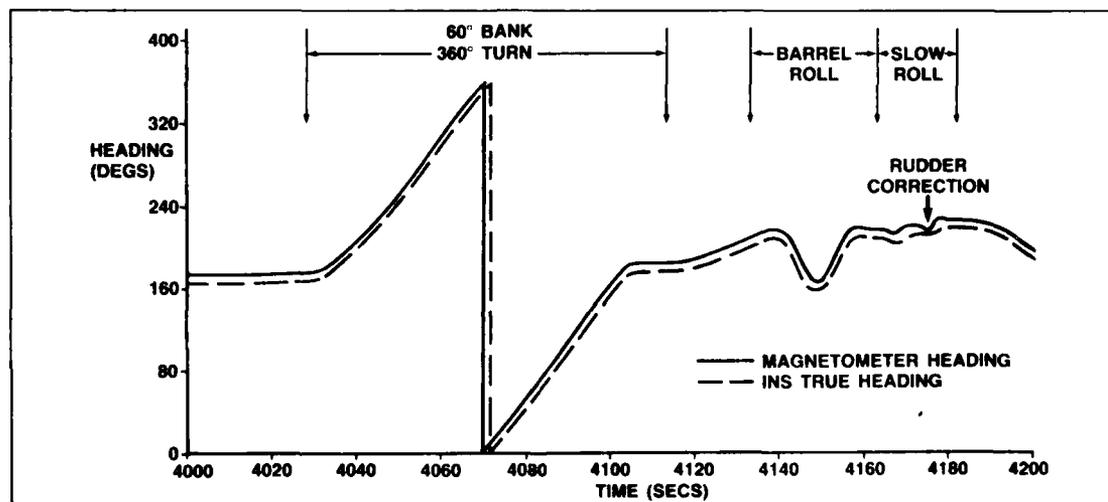


Figure 12 3 Axis Magnetometer Heading

The next stage of development would be to Kalman filter the basic magnetometer heading with the rate of change of heading resolved through attitude from the FCS gyroscopes. As the gyroscopes are near the centre of the aircraft, they will be unaffected by deflections of the fin and therefore the filtered value should show a marked attenuation of the errors seen in the roll.

NAVIGATION

The heading provided by the 3 axis magnetometer, together with the projection of True Air Speed onto the horizontal plane provide the basic elements for an Air Data, dead reckoning navigation system. Such type of navigation is susceptible to errors induced by changes in wind speed and direction and is therefore only used in emergency situations when all else has failed. However, over short periods of time the accumulation of navigation error may be sufficiently small to enable further exploitation in association with other techniques. One such technique is terrain profile matching which estimates position by correlating measured height of the terrain profile with stored height of the terrain being overflown. The technique works in conjunction with a dead reckoning navigation system which is needed to provide initial estimates of direction and horizontal displacement for each correlation step. Current studies are associated with Inertial Navigation but preliminary investigations by BAe Dynamics, Bristol Division based on flight data suggest that their TERPROM continuous terrain profile matching technique may be workable with Air Data navigation providing heading errors are within a degree or so. The comparatively high update of continuous profile matching enables the system to cope with changes in wind conditions. The position accuracy will be less than TERPROM aided IN but is expected to be better than an IN alone. A further advantage of this technique is that it can also provide accurate horizontal and vertical terrestrial velocities.

The attraction of continuous terrain profile matching is its autonomy with a high confidence afforded by virtually continuous, automatic position fixing unimpaired by weather conditions. To this can be added the advantage of using the associated height data base to estimate the height of ground ahead of the aircraft, without the emission of forward energy, again regardless of weather conditions, to provide the basic range data for a passive terrain following system.

Terrain profile matching of course is limited to flight over land for which there is a digital height data base. An alternative high accuracy navigation system is NAVSTAR which, being radio based, lacks the autonomy of terrain profile matching but at least it is Global in its application. This system essentially provides measurement of aircraft terrestrial position and velocity, but cannot provide heading or attitude data. Consequently, a cost effective solution could be to provide attitude and heading by the Synthetic Attitude Technique in conjunction with the three axis magnetometer. Air Data navigation could be used to sustain continuity during brief disruptions of NAVSTAR satellite communication either through jamming or obscuration of the antennas in dynamic manoeuvres, and would also assist in reducing the time to reacquire satellites.

DISCUSSION

The prime objective of the flight demonstration was to show that the concept of extracting attitude from the flight control system was fundamentally viable. Absolute accuracy was not a major consideration provided any inadequacy could be accounted for. As the flight results confirm the general performance is encouraging. There is some loss of performance in dynamic manoeuvres associated with high manoeuvre rates but not of a nature to invalidate the concept. As refinement of the performance would require substantial changes to the aircraft instrumentation recording standards, it was felt that before actively pursuing further development there should be a potential aircraft application, and here a further factor had come into play.

Aircraft attitude is important as a reference for the regulation of the aircraft's flight path. Before the introduction of inertial navigation, a climb or dive was set up by reference to the artificial horizon and rate of climb display and would take a succession of control readjustments before the required flight path was achieved. The inclusion of an inertial navigator made it possible to depict the aircraft's flight path relative to the outside world. Pilots find this display compelling as it gives a direct indication of how the aircraft is responding to control demands and thereby made establishing a flight path a much simpler task than hitherto.

The aircraft flight path display is based upon inertial navigation (IN) terrestrial velocities. IN systems are difficult to endow with a high degree of self fault detection and ground speed runaways can arise without any fault indication. Such a velocity increase would reduce the magnitude of displayed climb/dive angle. As the horizon display, also derived from the IN, would not be significantly affected by the runaway, the pilot may be unaware that an actual dive angle was much greater than that displayed and this could result in a hazardous flight condition. Consequently there is much concern to improve the integrity of the flight path display. Attempts to monitor IN ground speed from non IN sources do not appear to be sufficiently accurate and this has led to demands for a second IN. This second IN would also improve attitude integrity.

The coming generation of the INs are likely to employ Ring Laser Gyroscopes (RLG) not least because they offer advantages of reduced life cycle costs. As the RLG measures rotation rates they could be used to provide axis rotation rates for the FCS and therefore offset the cost of multiple IN equipments. Thus one might expect to find in multiple lanes of an FCS at least two lanes using the outputs from IN RLGs.

Looking further ahead the sensors for both IN and FCS requirement might be provided by the Integrated Sensor Package based on the RLG. The accuracy would be sufficient for IN requirements, the integrity and availability sufficient for FCS and therefore the integrity and availability of flight path angle and attitude data would be assured.

All these considerations have tended to detract from the initial interest in the Synthetic Attitude technique, and currently there is no projected application with the consequence that further development has been low key.

However, it has already been mentioned that continuous terrain profile matching and Navstar provide high accuracy terrestrial velocities and clearly therefore could provide an alternative source of flight path angle to an Inertial Navigator. This prospect offers many alternative future system applications for using Synthetic Attitude and the three axis magnetometer with either or both terrain profile matching and NAVSTAR, without an Inertial Navigator or to compliment an Inertial Navigator.

CONCLUSIONS

1. An investigation has been undertaken into the possibility of achieving a high integrity, high availability source of pitch and bank attitude estimates from the sensor data within a digital Active Control Technology flight control system. The outcome is a proposal to compensate for the relatively high drift rate of the ACT gyroscopes by improving the long term gravitational reference through correcting acceleration measurements for inertial accelerations derived from the ACT Air Data sensors and gyroscopes.
2. Flight demonstration of an experimental system on the digital ACT demonstration Jaguar aircraft has shown that the concept is viable. Accuracy is good in prolonged gentle manoeuvres and although bank errors were large in high dynamic manoeuvres, the errors were insignificant by completion of the manoeuvre. Subsequent theoretical studies based on flight data have shown that the bank error can be reduced.
3. Further development on manned aircraft has been overshadowed by the additional need for high integrity flight path data driving solutions toward multiple inertial navigation systems.
4. In the context of low cost systems, it should be possible to extend the role of this source of attitude by association with other technical developments to arrive at an alternative to the inertial navigator as a source of accurate terrestrial position and velocity and therefore flight path angle.

Firstly, attitude can be used to extract heading from a three axis magnetometer which can be smoothed by the FCS gyroscopes. This heading, in conjunction with True Air Speed could provide an Air Data navigation capable of being aided by continuous terrain profile matching to yield high accuracy navigation and terrestrial velocity.

Alternatively, for operation over sea or land not covered by a digital height data base, terrestrial positions and velocity can be provided by NAVSTAR satellite navigation with attitude and heading provided by the Synthetic Attitude Technique and the three axis magnetometer. Air Data Navigation could be used to interpolate during brief interruption of satellite communication during severe manoeuvres or periods of jamming and also assist in minimising reacquisition time of a satellite.

Whether or not such types of system would ever be capable of displacing the Inertial Navigator is perhaps speculative at this time. However, at the very least they may be capable of complementing an IN to improve flight path integrity etc without the need for a second IN.

THE IMPACT OF VLSI ON GUIDANCE AND CONTROL SYSTEM DESIGN

by
DOM PRICE
 Flight Automation Research Laboratory
 GEC Avonics Ltd
 Airport Works
 Rochester Kent ME1 2XX
 UK

SUMMARY

The potential of Very Large Scale Integrated Circuit technology to reduce cost, including Life Cycle Cost, of Avionics systems is examined. Trends in methods of silicon realisation and in system complexity are related to the necessary advances in methods of system design. A generic system design approach is outlined, which emphasises the disciplines which are likely to enable cost-effective semi-custom circuit design to become viable.

1. HISTORICAL INTRODUCTION

In the 1960's the aerospace industry defended its public image, always under attack for increasing expenditure, by holding up the non-stick frying pan. Thank to aerospace technology the housewife could deflect her husband's anger at paying taxes by offering him perfectly-cooked, never-burnt fried food. If in retrospect we look for commercial and domestic benefits of aerospace technology the largest benefit must be seen in the collaboration between the aerospace industry and the semiconductor industry in the development of the ubiquitous capable and reliable integrated circuit. For which other product can we look back on two or three decades in which value for money has increased at such a pace?

Today we see the rapid development of very high performance integrated circuits to solve problems of signal processing and image analysis stimulated by the need for smart weapon systems, with the commercial drives of the Information Technology industry forcing the pace just as positively. More and more powerful microprocessors find their way into avionics, a tiny minority sponsored by the aerospace industry. Certainly the microprocessor has enabled a considerable advance in processing power per unit cost and a large step forward in reliability. It is the object of this paper to assess what further steps in cost effectiveness might be obtained through the use of the VLSI, and what the two industries need to do to make this possible.

2. AREAS OF SYSTEM COST AMENABLE TO REDUCTION

Let us first distinguish between first cost and cost of ownership. The cost of an integrated circuit depends on the size of the market into which it is sold, the maturity of the product, competition and the clout of the Buying Department. Cost of ownership relates to reliability, ease of fault diagnosis and repair, and less directly to the systems heat dissipation, power requirements, connector requirements, weight and bulk.

Trade-offs between first cost and cost of ownership are intrinsically difficult as the system supplier has little control over the detailed maintenance and repair tactics employed by his customer, but in principle all these areas are amenable to improvement by further application of VLSI. Apart from the highest speed areas of a circuit, dissipation and area of silicon real estate are markedly controlled by the need to provide drivers to transfer signals from chip to chip, and further levels of integration of circuits can dramatically lower system power levels. A more logical division of systems into functional blocks can ease the diagnosis of failures leading to a significant reduction in the proportion of integrated circuits incorrectly removed and a longer life for the boards on which they are mounted.

Additionally a more logical partitioning of systems into integrated circuits facilitates the incorporation of built-in self-test facilities into the chips. A standardised interface between circuits with spare capacity enable interrogation of the BIST function without additional pin-outs and drivers. Since the reliability of an integrated circuit is largely independent of its circuit complexity, this is achieved without penalty in MTBD, and the failure of BIST circuitry leads to the removal of the correct chip.

However such ideal solutions are unlikely to be carried through into practices unless and until the avionic industry can enforce standards and can afford the first cost of chips to its own specification. Its ability to carry through such apolicy at today's juncture is doubtful indeed; this is because the first cost of such devices would be dominated by design cost.

Trends in Design Cost of Integrated Circuits

In 1980 Frank Micheletti of Rockwell illustrated the trends in design cost over the decade; the cost of IC design rose linearly as the number of gates increased. Although we could see promise for decline in per gate design costs with advanced CAD by

impression, for full custom circuits, is that \$100 per gate still holds certainly for microprocessors and other circuits where a high degree of geometrical repetition does not dominate. For such circuits where detail layout is important to fulfill performance and yield goals there is little sign of a marked fall in design costs per circuit unit. However where a design is required that does not need optimal packaging to obtain performance or yield, some different approaches may be made which lower design cost, principally the Gate Array and the Cell-Based System, to produce Semi-Custom designs. Semi-custom design however should not be defined in terms of techniques such as the above because it limits our concepts of system design; gate arrays are only the on-silicon analogy of design by connecting up individual transistors, and cell based design can take us back to SSI technology fitting together flip flops, D types or perhaps registers. A better concept of semi-custom design is systems design where units of circuit in which the system designer has experience and confidence are interconnected on silicon to meet a requirement. The circuit units could conceivably be similar to the common devices he would look for in a pcb design; registers, buffers, decoders or even microprocessors leading to what is commonly termed a macrocell approach. The silicon vendor is thus potentially able to 'shrink' mature design into a finer geometry process, and by combining them meet a system requirement in a fraction of the previous number of devices. Such combined systems however are likely to have a more limited market, will need system knowledge to select and design (and equally to specify tests for) and except in limited areas offer a poor return for the design resources of a component manufacturer.

However as a joint design exercise for the systems manufacturer and the silicon designer the approach deserves examination if the design process can be mechanised in an error-free manner, the macrocells are proved by repetitive manufacture and use, and the overall circuit with test access and test routines can be confidently simulated by the customer (since 'breadboarding' of designs is impossible).

'Software Preservation'

There is one further aspect worth considering, which is that of software. The major of avionic systems are software programmable, not just because a microprocessor is currently cheaper than a custom logic circuit, but because a significant number of design parameters cannot be finalised when the design must be committed. These include airframe derivatives, weapon constants, sensor properties, symbology and other man-machine interface parameters. Flexibility in design is thus an inherent need, not a side benefit of digital technology. Whereas to an extent software may be separately developed on a host machine and down loaded, to use a host machine for real time software commissioning in conjunction with breadboard peripherals is expensive. Hence the tendency to demand a breadboard for real time software proving using available microprocessors. On design completion however the breadboard needs to be 'shrunk' into a silicon solution. If such a macrocell approach were available and the major components had been selected for process compatibility, then instead of rewriting the software to suit the compact hardware solution we could simply shrink the circuitry into macrochips and we could perhaps preserve our software investment intact. A further benefit is that test routines would not have to be reworked.

3. STEPS NEEDED TO ACHIEVE COST EFFECTIVE SEMI-CUSTOM DESIGN

3.1. Workstations

A whole new business has recently mushroomed into existence in the supply of engineering workstations. These enable engineers to design semicustom circuits without continuous access to a mainframe computer and hence provide a cost-effective design tool.

Much of the design work can be independent of the design rules of a particular silicon process. When the design reaches the point at which the designer needs to incorporate the particular rules of a chosen or candidate process the workstation permits him to do so, the dominant parameters having been supplied by the process manufacturer to the designer of the workstation. By an extension of such a facility it is likely that the workstation could access the silicon house's standard cell library and progressively allow design to take place at a higher level, still allowing the workstation user the facility to design his own cells from the basic design rules and integrate his design at a mixture of levels.

3.2. Test Methods

A further natural extension to the workstation principle to the inclusion of access to cell libraries could be to add a library facility of test strategies. Besides making available historic data on proved test approaches to cells it could prove feasible to adapt a cell of proved testability to solve a particular logic problem rather than adopt a logic solution and subsequently attempt to wrap a test strategy around a less amenable logic array.

3.3. Simulators

Inherent in proving a design before committing it to prototype manufacture is the use of a logic simulator. In a host machine each step in the propagation of inputs through the logic solution is simulated in software, albeit at a minute rate compared to the reaction time of the system to be realised. Such simulators are powerful analytic tools in showing up weaknesses in the design. However the slow speed operation can be an embarrassment, and when it comes to simulating a complex system with software embedded

the slow processing holds up the design process. Two steps in alleviating this problem are the use of powerful parallel processors in the host machine and/or the use of real time hardware as part of the simulator.

3.4. Floor Plan Design

The top level layout of major cells in a VLSI circuit is conventionally termed floor plan design. This step is crucial to economic design and is a highly heuristic process in which a limited number of individuals excel. If we are to open up systems design in silicon to a greater number of individuals this is an area worthy of attention along with partitioning a system into a workable set of chips. It is an area of design activity which is believed amenable to a Knowledge Based System approach where a number of straightforward rules of play need to be combined with a strategic feel. It is likely to be amenable to the synthesis of rules by example in the way that the rules of chess end games have been developed from interaction of knowledge engineers with chess masters and the encoding of examples into generic rules by an Expert Systems.

3.5. Speeding up the Design Process

The impression that tends to be given of the design process in silicon is of a steady top down process, mapping each level of design and verification exactly on to the next process by employing a set of layered design tools in sequence. This before the introduction of workstations is probably the ideal to which designers aspired, as the entry of design data from one layer to the next tended to be a manual process, error-prone, and one no designer wanted to carry out more than once. However, this is not the way in which practical designers proceed in other fields. Given a highly integrated set of tools with minimum human intervention in passing from one layer to the next, and a comprehensive set of library information accessible at all layers the designer will revert to his normal mode of iterating rapidly up and down the process, trading off floor plan estimates with testability criteria and shaping his design towards an optimum in a way that a strict top down mapping process is unlikely to permit.

There are other reasons which drive the design process towards an interactive evolutionary strategy. The systems designer wishes to keep his options open as long as possible not only on which silicon process to select but which vendor to place an order with. He wishes to go through the design process to the point at which he can trade off more than one silicon technology and to have first cut design data to offer a prospective vendor to obtain a quotation. The prospective vendors may wish to suggest changes and he will have to rework his framework to assess their impact. All these desirable attributes are feasible if the design suite is integrated and has access to the appropriate cell, design rule, testability and layout library functions.

3.6. Recipes for Successful Implementation

Such a plan as described depends on the quality of the tools, libraries and design rules. It is axiomatic that the quality of software depends largely on the size of the user base, the willingness of users to report defects and of generators to eliminate them. Broadening the user base as described should have the requisite effect of detecting and reporting errors. An efficient software update service, possibly over data links, should prove a more effective control than amending catalogues. A more effective control by the silicon vendor in correcting process defects and amending design rules will be needed to maintain a qualified process tolerant to a wide range of designers.

Experience however indicates that the greatest single hindrance to right-first-time design lies in errors, inconsistencies and misunderstandings at the initial specification stage. This is also the area in which VLSI design aids are probably the least advanced. However recent trends in logic programming languages and other declarative languages suggests that the AI community may again have some solutions to offer.

3.7. A Final Suggestion

If this general approach becomes feasible in the years to come it would also become feasible to standardise test interfaces, introsystem bus standards etc to the point that in a similar way to the adoption of MIL STD 1553B for data exchange between systems and the availability of standard terminal chips, we might very well find in the cell libraries cells which are qualified to MIL Standards and available in more than one technology.

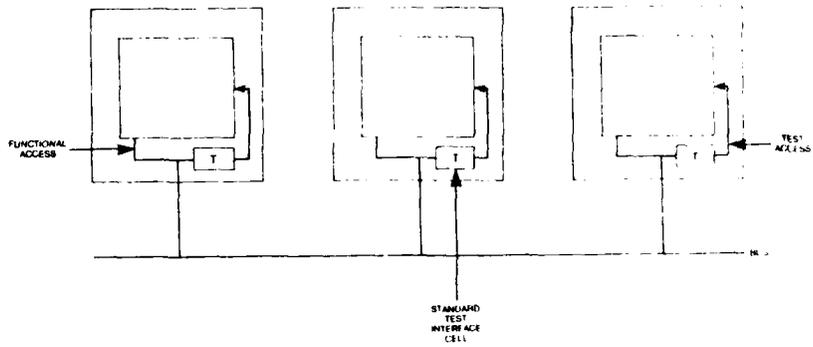


Figure 1

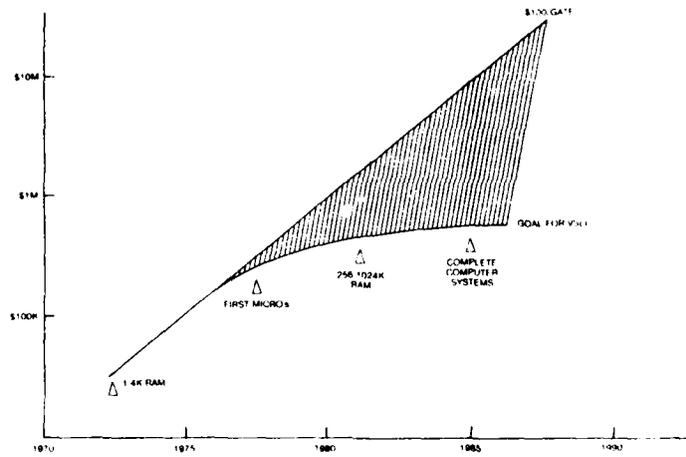


Figure 2

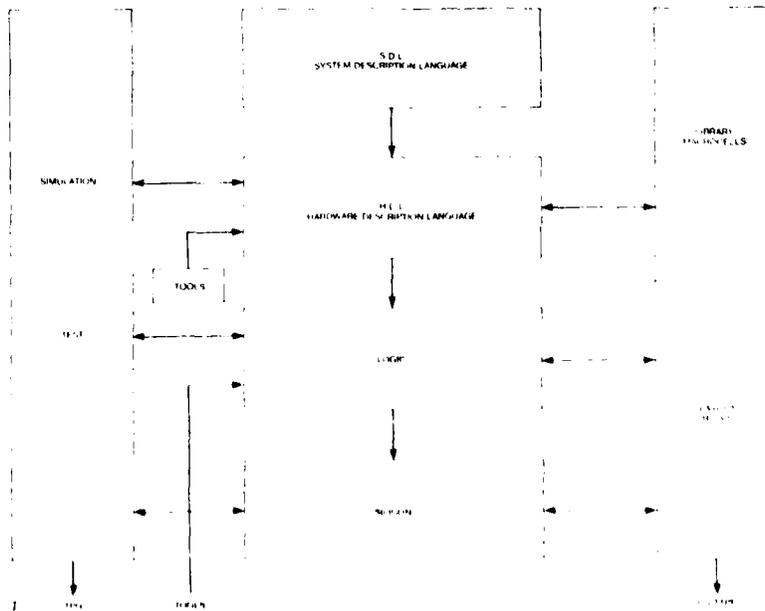


Figure 3

REUSABLE SOFTWARE - A CONCEPT FOR COST REDUCTION

Christine M. Anderson
Marlow Henne*

Air Force Armament Laboratory, AFATL/DLCM
Eglin Air Force Base, Florida 32542
U.S.A.

Summary

The cost of military computer systems is increasing rapidly. It is alarming to consider that not only the total cost of military computer systems is increasing, but the percentage of the total cost attributed to software is also increasing. Several underlying causes for this increase are discussed. While the new United States (US) Department of Defense (DoD) standard high order language, Ada, will significantly help to reduce the software cost growth, other solutions must also be sought. Reusable software component technology and associated parts composition systems are presented as possible solutions. Recommendations for future research are provided.

1. Software Cost Growth

Much growth in the power and sophistication of US defense systems is due to the extensive application of computers. Almost every defense system fielded today contains computers whose software performs mission-critical functions [1]. Most of these are embedded computers. An embedded computer can be thought of simply as an integral component of a larger system. Defense systems or subsystems using embedded computers include sensors, electronic warfare systems, weapon system control, communications, command and control, navigation, and target acquisition.

The cumulative inventory of DoD embedded computers is projected to grow from 10,000 in 1980 to nearly 260,000 by the end of the decade. While hardware quantity is increasing rapidly, the percentage of the total dollar attributed to software will increase from 65 percent of the total in 1980 to 85 percent of the total by 1990. This translates to a staggering projection of \$32 billion (absolute dollars) for annual DoD embedded software expenditures in 1990 compared to \$5.9 billion for DoD embedded computer hardware [2].

There are several underlying causes for the increase in software cost over the past decade and the projections cited above. A prime reason is the large variety of programming languages that has evolved. By 1974, there were over 400 languages, dialects and subsets of which few supported modern methodologies for structured program development, making maintenance (accounting for 65% of the software life cycle cost [3]) a nightmare [4].

Another contributing factor to the increased expenditure in DoD embedded computer software is inherent in software itself--its flexibility. DoD has been increasingly exploiting software's flexibility in developing modern weapon systems. As reported by the USAF Scientific Advisory Board, software "can embody and implement abstract operational concepts; it has no manufacturing cycle or cost; it can be modified quicker and cheaper; it does not wear out; and it can incorporate new features and functions in an evolutionary fashion without major investment in new systems and hardware" [5].

The Air Force F-111 program illustrates this point. The table below compares similar capabilities (additional offset aim pointer and updated weapon ballistics) implemented through hardware on the F-111 A/E and in software on the F-111 D/F. Given an existing software support facility, the savings due to making the changes via software rather than hardware have ratios of about 50:1 in cost and 3:1 in time [6].

| <u>Modification</u> | <u>Via Hardware</u> | <u>Via Software</u> | <u>Cost/Time Ratios</u> |
|---------------------|---------------------|---------------------|-------------------------|
| #1 | \$5.28M/42 mo | \$0.10M/16 mo | 52.8:1/2.6:1 |
| #2 | \$1.05M/36 mo | \$0.02M/10 mo | 52.5:1/3.6:1 |
| #3 | \$8.00M/78 mo | \$0.02M/15 mo | 400:1/5.2:1 |

Another cost contributor is the fact that software development is labor-intensive. Typically each line of a computer program is written by hand. Unfortunately, while the demand for software is increasing, the number of qualified personnel is not increasing as rapidly. The Air Force Scientific Advisory Board reported that there is a 4% annual increase in qualified software personnel, a 4% annual increase in software productivity (based on current methods) and a widening gap based on a 12% annual increased demand for new computer software [5].

In more absolute terms, the shortfall between supply and demand, currently measures 50,000-100,000 programmers, and may rise to 1.2 million by 1990 if remedial measures are not taken [7].

*Mr Henne is now at Harris Corporation, GISS Software Operations, Melbourne, Florida 32901

Another cost contributor is the problem of building reliable software. Instances of software reliability problems include false alerts for the North American Defense (NORAD) system, space shuttle's on-pad launch delay, and test missiles (and even airplanes) hitting mountains they were programmed to fly over [5]. The criticality of the reliability problem is made clear by C.A.R. Hoare's warning: "The next rocket to go astray as a result of a programming language error may not be an exploratory space rocket on a harmless trip to Venus. It may be a nuclear warhead exploding over one of our own cities" [8]. Developing reliable mission critical software where errors can translate into life or death situations is a time consuming and costly process.

To briefly summarize, the reason for the tremendous cost growth in defense embedded computer software is multifaceted. Contributing factors include: the large number of existing computer languages currently being used; the increased exploitation of software's flexibility resulting in its increased utilization; the labor-intensive nature of software resulting in a shortage of qualified software engineers; and the complexity associated with meeting the reliability requirements of mission critical software. Solutions to all but the second factor can be, and currently are being sought. The second factor is really a reflection of our "age of information" rather than a problem.

One of DoD's more notable technological and managerial solutions is the development and standardization of one high order language for mission-critical computer systems--Ada. While Ada has a tremendous potential to reduce costs in all phases of the software life cycle, Ada alone is not a panacea for DoD's software problems. We cannot hope to meet the software needs of the 1990s and beyond if we continue to develop Ada software on a line-by-line basis. One approach is to develop the concept of software reusability using Ada as the cornerstone.

2. Software Componentry

The reuse of software components has the potential of reducing the shortfall of required software engineers, while increasing software reliability.

There is a great deal of interest in software component technology. The following discussion represents a collage of thinking on the subject that has been generated over the past several years.

"Having reusable software available can significantly reduce system development time. The more software that can be obtained off-the-shelf, the less new software that must be created. The risks involved are thus reduced, since off-the-shelf software should already have been well tested and debugged. Reduced time and risk enhance the probability that a project will be completed on time and within budget...Such benefits constitute a major resource savings, ultimately reducing the required DoD software investment." - Commander J. Cooper, "Increased Software Transferability Dependent on Standardization Efforts," Defense Management Journal, October 1975.

"Software reuse saves development time and money, and field proven software is more reliable." - Strategy for a DoD Software Initiative, 1 October 1982.

"The Introduction of reusable software components can significantly relieve the resource demands thus assuring continued responsiveness to new threats through the introduction of new or enhanced weapon systems". - Report of the DoD Joint Service Task Force on Software Problems, 30 July 1982.

Recommend the Air Force "Initiate a set of formal laboratory programs to define and exploit opportunities for software standardization through reusable software packages in selected application areas." - Dr Edwin B. Stear, former Chief Scientist of the Air Force in a Briefing to General Marsh, AFSC Commander, 14 March 1983.

"Achieving reusability in mission software represents a good opportunity for dramatic productivity gains since using existing software in lieu of new software not only saves money but also saves documentation costs and test costs. And since the software has already been verified, it increases the quality of the new system." - The Army Science Board 1983 Summer Study on Acquiring Army Software.

2.1 A Hardware Analogy

There is nothing unique about component technology: it has existed for years in the digital electronics world. The uniqueness emerges when we apply this technology to software. To a great extent, it can be said that today's software design process has not progressed much beyond the early digital design methodology in the digital hardware industry.

In the late 1950's and early 1960's low level components, such as vacuum tubes, transistors, resistors, and capacitors, existed for the design engineer to use for his circuit design. A strong analogy can be drawn between these and the assembly language statements used by the software engineers of today. Each time a particular kind of circuit, such as a gate or flip-flop was needed, the engineer had to select the type of circuit and the components' values which would be used. When large quantities of a particular kind of circuit were needed on a project, the engineer could reuse his design

for that project. The design of each circuit for large projects was such a chore, individual engineers kept earlier designs and tried to borrow from them on new projects whenever possible. This also led to published design compendia of circuits much as may be found today in collected algorithms and published math or utility routines.

Enterprising companies realized that significant markets existed for certain low level digital building blocks and began to develop circuit cards which contained individual gates, flip-flops, etc. These circuit cards (modules) made a significant contribution to the design of systems. By using these ready made modules, the designer was freed to do more productive system design. While these modules were helpful, a significant problem still existed when trying to use modules from different suppliers. Each supplier chose his own voltage levels for logic "one" and "zero", and other interface details were often different.

This level of modularity is close to that which we are approaching today in software. We have standard math and utility routines which exist in libraries but little else is available for general reuse.

Once the advantages of reusing hardware designs became evident, the next step was to improve the production technology for hardware modules. As integrated circuits (ICs) became possible, computer and general digital components represented the vast majority of units built. This, of course, was due to the modular nature of digital designs with their many common components and to the design methodology which had already begun supporting reusable module designs. ICs enabled entire flip-flops and other circuits, which had previously taken an entire circuit card, to be fabricated on a single IC. As production technology improved, more circuits were placed on individual ICs to form intermediate modules such as counters, adders, etc. Today, we find Very Large Scale Integrated (VLSI) circuits which represent large portions of the overall system. Since these circuits are so large in scope, most systems may not use more than one of a given circuit. Even though we may only need one circuit of a given kind, production techniques and reusability have reduced the cost of these complicated circuits to a level where they are competitive to use for a variety of applications. A good example of this is the microcomputer. Today we find full-scale digital computers, on a single IC, performing timing tasks for small appliances which previously used analog circuits or mechanical timers.

If the current trend continues toward software module production and reuse, we should see the size and use of common software modules expand rapidly, paralleling the development of hardware modules.

Thus, an analogy can be made between the inventory of software components and the inventory of prefabricated circuits available from semiconductor manufacturers. The software developer resembles the computer designer, who determines the gross system structure and the interconnections between circuits but relies on the prefabricated components for low-level operations. Over the past few years, larger components have been developed. A similar technology is needed in the software world [9].

2.2 Software Componentry Methodology

Past critics have maintained that the software reusability concept has not flourished because individual concrete programs are too specialized to be reused. However, with the advent of Ada, we now have the means to more easily construct software components at a more abstract level. Ada's package feature will permit the creation of software packages analogous to sealed hardware components that consist of an external interface or specification and an internal body which the user cannot alter. Ada's generic feature extends the package concept to include a parameterization facility for tailoring packages to particular needs.

Additionally, Ada's strong typing imposes constraints on module interconnections and allows consistency between formal parameters of module definitions and actual parameters of module invocations to be enforced at compile time [10]. These features provided by Ada for reusable software components are richer than those of its predecessors.

Studies investigating methodologies for applying Ada to develop reusable software components are only now being initiated. Thus the technique involved in developing generic packages is not well defined and almost no implementation experience exists. A recent Air Force study identified criteria which impact the reusability of software. Chief among these are application independence, modularity, simplicity and code self-descriptiveness [11].

Application independence can best be achieved via generalized data structures such as parametrically described arrays; minimal use of machine dependent constructs such as machine language code, microcode and specific I/O features; and implementation of well chosen common functions.

The components should be encapsulated in such a way that their external usage is completely defined by an interface specification, which is physically distinguishable from the implementation portion. This interface should be as firm and well defined as that for the hardware interconnection to an integrated circuit. Further, components should be orthogonal. This means that, unless contraindicated by their interface descriptions, they may be used in each other's presence [12].

Simplicity in both design and implementation, will facilitate program understanding and modification. The quantitative counts (number of operators, operands, nested control structures, nested data structures, executable statements, statement labels, decision points, parameters, etc.) will determine to a great extent how simple or complex the source code is.

Component source code should be as self-descriptive as possible. One approach to achieving this goal is to embed compilable program design language (PDL) in the source code, thus insuring up-to-date cross-correlation between design, code and documentation. By using valid Ada procedure names and declarations in addition to commentary in the PDL design, rigorous checking of the PDL can be performed.

In order to encourage use of software components once they are implemented, a systematic approach to accessing and combining particular instances of components must be pursued. This approach focuses on parts composition system technology.

3. Parts Composition Technology

A parts composition system supports the building, testing, and optimizing of programs using reusable components. This system must include, at a minimum, cataloging and retrieval facilities, a language to compose parts, a warehouse of parts, and an editing and testing facility. The Japanese have reportedly achieved up to an 85% reuse rate in their software factories by using these currently available information retrieval techniques [13]. Toshiba's Fuchu Works Software Factory, specializing in real time applications, averages 2870 instructions per programmer per month [14], compared to a U.S. software productivity rate of 75 to 280 lines per programmer per month [15]. Toshiba's productivity is due in large part to software reuse. A mature parts composition system will include thousands of software parts available in a common environment. In theory, a software engineer can attempt to combine any two available parts so the system must provide robust mechanisms to insure reliable and meaningful parts composition [16].

Parts composition systems may range from manual and semiautomated software parts catalogs to more advanced automated systems, systems that may even include artificial intelligence (AI) (i.e., expert systems).

An expert system is a man-machine system with specialized problem solving expertise. The first generation of AI focused on defining a general mechanism of intelligence for expert systems. The current perspective holds that the true power of the expert system comes from the knowledge it possesses, not from the particular formalisms and inference schemes it employs [17]. Thus, it is essential to capture the knowledge of the application domain to be modeled.

While various domains of knowledge are being studied and common functions extracted for later component implementation, parallel investigations of methods to organize, index, describe, and reference software components must also be pursued. Further in conjunction with all of these activities, studies aimed at producing more advanced user friendly systems that change data and algorithmic representations into code, that is, software generator systems must continue.

The following example describes one scenario of a user interacting with a knowledge-based parts composition system. Assume the user is interested in locating a guidance algorithm for a particular armament electronics (armonics) application. He asks the system for retrieval of all generic classes of guidance software that meet his requirements. The system may actually solicit (via leading questions) the requirements from him. After reviewing these generic classes, he asks the system to retrieve more detailed descriptions of certain components. Following the examination of these, a more detailed review of the specifications associated with a select few components is performed. Finally, the actual component code is examined. This retrieval mechanism can be made increasingly intelligent by providing facilities for querying the user for additional information if no component is found, by searching for related components or by custom tailoring components to match the user's request.

This example is still at a fairly low level of automation. A more advanced parts composition system would allow the user to describe (in a user oriented high level specification language) an entire subsystem's requirements (e.g., autopilot for a particular type of weapon). The system would query the user concerning critical aspects of the subsystem and then proceed to retrieve, customize and compose the necessary software.

There are several parts composition systems commercially available that offer varying degrees of aid to the user. Thus far, none have been applied to the development of weapon system software. An evaluation of these systems should be performed.

4. A Related Effort

The US Air Force Armament Laboratory has just initiated a program that addresses software component technology and supporting parts composition systems. The program, Common Ada Missile Packages (CAMP), features two related study efforts: a commonality study and a parts composition study. The objective of the commonality study is to investigate the feasibility of developing reusable Ada components for armonics systems. The

approach is to examine existing missile software and/or associated documentation in order to identify candidate common functions for component implementation. The second study, to be performed concurrently, features an investigation of current parts composition system technology and recommendations regarding the most practical approaches for achieving both near-term and long-range benefits. Based on the results of the studies, a follow-on implementation phase will commence aimed at developing a parts composition system and associated reusable software harmonics components.

5. Conclusion

In closing, it should be stressed that software componentry will not evolve quickly or cheaply. Both mental and organizational road blocks must be overcome. However, the technology holds such a tremendous potential for slowing the cost growth in DoD mission-critical software, further research is imperative. This research should be aimed at joining parts composition technologists, who are often from academia, with DoD mission-critical specialists in order to achieve a fruitful blending of composition techniques and DoD knowledge domains.

References

1. Martin E.W., "The Context of STARS," IEEE Computer, November 1983.
2. Electronics Industries Association Government Division, "DoD Digital Data Processing Study - A Ten-year Forecast," October 1980.
3. Grove H. M., "DoD Policy for Acquisition of Embedded Computer Resources," Concepts, The Journal of Defense Systems Acquisition Management, Autumn 1982 Volume 5, Number 4.
4. Deutsch R., "JOVIAL: The Air Force Software Solution in the Years Before Ada," Defense Electronics, October 1982.
5. USAF Scientific Advisory Board, "Report on the High Cost and Risk of Mission-Critical Software," December 1983.
6. DoD, "Report of the DoD Joint Service Task Force on Software Problems," July 1982.
7. Boehm B.W., Standish, T.A., "Software Technology in the 1990's," Appendix to Software Initiative Plan, October 1982.
8. Hoare C.A.R., "The Emperor's Old Clothes," Communications of the ACM, Volume 24, Number 2, February 1981.
9. Wasserman A.I., Gutz S., "The Future of Programming," Communications of the ACM, Volume 25, Number 3, March 1982.
10. Wegner P., "Varieties of Reusability," Workshop on Reusability in Programming Proceedings, Sponsored by ITT Programming, Stratford, Conn., 7-9 September 1983.
11. Presson P.E., et al., "Software Interoperability and Reusability," Boeing Aerospace Company under contract to Rome Air Development Center, Griffiss AFB, NY, NY, RADC-TR-83-174, July 1983.
12. Spector D., "Language Features to Support Reusability," SIGPLAN Notices, ACM, Volume 18, Number 9, September 1983.
13. McNamara D., "Japanese Software Factories," NSIA Conference, Arlington, VA, May 1984.
14. Kim K. H., "A Look at Japan's Development of Software Engineering Technology," IEEE Computer, May 1983.
15. Zekowitz M. V., Yeh R. T., Hamlet R. G., Gannon J. D., Basili V. R., "Software Engineering Practices in the US and Japan," IEEE Computer, June 1984.
16. Rice J.R., Schwetman H., "Interface Issues in a Software Parts Technology," Workshop on Reusability in Programming Proceedings, sponsored by ITT Programming, Stratford, Conn., 7-9 September 1983.
17. Hayes-Roth F., Waterman D.A., Lenat D.B., Building Expert Systems, Addison-Wesley Publishing Company, Inc., Reading, Mass, 1983.
18. Berard E.V., "Ada Education-A Moving Target," Defense Science & Electronics, May 1984.
19. Dolotta T.A., et al., Data Processing in 1980-1985, John Wiley & Sons, NY, NY, 1976.
20. Boehm B.W., Software Engineering Economics, Prentice-Hall, Englewood-Cliffs, NJ, 1981.

21. Standish T.A., "Software Reuse," Workshop on Reusability in Programming Proceeding, Sponsored by ITT Programming, Stratford, Conn., 7-9 September 1983.

22. Bunyard J.M., Coward J.M. "Today's Risks in Software Development-Can They be Significantly Reduced?" Concepts, The Journal of Defense Systems Acquisition Management, Volume 5, Number 4, Autumn 1982.

23. "Missing Computer Software: A Bottleneck Slows New Applications, Spawns a Booming New Industry," Business Week, September 1, 1980.

MAINTAINING CONSISTENCY
AND ACCOMMODATING NETWORK REPAIR
IN A SELF-REGENERATIVE
DISTRIBUTED DATABASE MANAGEMENT SYSTEM
(A discussion of the means and costs)

M. REILLY, P. R. TILLMAN,
R. CATT, R. B. MOORE

SOFTWARE SCIENCES LIMITED
THOMSON HOUSE
292 FARNBOROUGH ROAD
FARNBOROUGH
HAMPSHIRE
GU14 7NU
UK

ADDAM is a database management system that maintains a real-time database which is partitioned and replicated amongst a number of computers connected in a local area network. With a centralised or quasi-distributed database, damage to particular components causes effective loss of the whole system due to loss of data. ADDAM can tolerate substantial damage to the network with little degradation of performance or loss of data. In this paper we describe some of the protocols used by ADDAM to maintain data consistency. Our work shows that it is possible to provide a very high degree of data reliability within a resilient real-time distributed database. However, the cost is quite large, both in terms of the amount of computer resources consumed by the database software and the number of inter-node messages generated on the network.

1. INTRODUCTION TO ADDAM (The A.R.E. Distributed Database Management System)

1.1 INTRODUCTION

ADDAM is a distributed database manager that maintains a real-time database which is partitioned and replicated over a local area network. ADDAM is part of a larger programme of work funded by the United Kingdom Ministry of Defence and carried out jointly by Software Sciences Limited and the Admiralty Research Establishment (Portsmouth). The network on which ADDAM runs is the ASWE Serial Highway [1]: this network has a very low error rate and supports a true broadcast facility.

ADDAM is characterised by the following features:

- a) data types partitioned into units of replication ('pages');
- b) replications of data partitions automatically generated for survivability and local access;
- c) synchronised updates for multiple data copies;
- d) concurrent update protection without record locking;
- e) consistency maintenance between multiple data copies under conditions of network damage;
- f) supports relational database model;
- g) user interface supports access to relational data model without reference to data location;
- h) choice of update protocols - reliability or performance;
- i) load balancing.

The background to ADDAM, together with a description of the network on which it runs, is given in [2].

1.2 PARTITIONING AND REPLICATION

The database is divided into a number of partitions (pages) [Figure 1]; a set of pages is allocated for the storage of all records of a data type. Where more than one page is allocated, the schema identifies a partitioning attribute for the data type, together with the range of attribute values for each page in the set. By adopting a suitable replication strategy, the two major requirements for the distributed database manager will be accomplished using suitable protocols i.e.

- a) survivability of data
- b) efficient local access to data subsets.

One of the key requirements of the replicated database is the need for synchronization between the pages holding the replicated data. This is achieved in ADDAM by using a master/slave arrangement. At any moment, one of the replications of each page acts as a master copy and the remainder act as slaves [Figure 2]. The master copy of the page is called the owner; the slave copies are referred to as subscriber copies.

An owner is not a privileged node on the network; ownership of a particular page may change from time to time as the processing load on the database varies. ADDAM attempts to transfer ownership to the node making most use of the page; the mechanism for deciding which node should own the page is called an election. In general, a node will be the owner of some pages and a subscriber to others and it will normally only have copies of a subset of the pages.

Two other techniques are used by ADDAM to optimise page distribution around the network:

- a) On each node ADDAM attempts to localise pages required for sequential access.
- b) ADDAM also ensures that sufficient copies of each page are maintained for survivability.

1.3 CONTROL PROTOCOLS

The database is maintained internally by a number of protocols known collectively as the control protocols. Their purpose is to maintain synchronisation and consistency among the multiple copies of the database, to ensure that the number of replications required for each page actually exist and to ensure that each page number has one, and only one, owner.

Update Protocols - Performance and Reliability

ADDAM's interfaces queue update requests to maintain single threaded user access on each node. The unit of update, the segment, is chosen to allow a record consisting of multiple segments to have updates applied to different segments concurrently.

The reliability protocol provides a reliable mechanism for the propagation of database changes to all copies of the relevant pages. It is used to support a number of application facilities including create and delete record and the reliability option of update. The reliability protocol has a handshake between originator and owner and between owner and subscriber. The reliability protocol also channels all database changes through the owner in order to achieve concurrent update detection and prevention.

ADDAM avoids 'double update' problems by rejecting reliability updates that do not have the latest version number in the segment header. This forces the application program to read before writing and to retry the whole transaction if a failure indication is returned by ADDAM.

The performance protocol is suitable for data updates that are both independent of the previous value of the data, and are applied at frequent intervals. This protocol dispenses with the handshakes used by the reliability protocol and with the channelling of updates through the owner at a cost of slightly lower reliability. The benefits are significantly reduced highway traffic, shorter protocol path lengths, and no waiting for acknowledgements.

The loss of a node may result in the loss of the owner or subscriber copy of one or more pages of data. A loss of ownership obstructs the passage of the reliability protocol which routes updates to data through the owner and thereafter to the subscribers. There are regular checks made by ADDAM that combine subscriber notification to the owner with monitoring of the number of owners that exist for each page. If the number of owners of a page is zero, an election is held between all subscribers, to elect a new owner. Bids are made on behalf of each subscriber which are evaluated in each copy of the election program. In theory all subscribers will come to the same conclusion about which copy of the page will become owner.

In practice there is always a possibility that the election evaluation program in one or more nodes does not receive all the bids and chooses owners differently. This can lead to a page having two owners. However, dual ownership can be recognised by ADDAM which forces another ownership election, disowning all current owners.

If a subscriber copy is lost, bringing the number of copies of a page below a schema-declared minimum, a new subscriber is selected from the remaining nodes. The owner broadcasts a "coerced subscriber request" message. This message is received by all nodes that do not hold a copy of the relevant page. Any node that is able to take a copy of the page sends a "subscriber enrolment offer" message to the owner. The owner accepts however many offers it requires, and rejects the remainder. The nodes whose offers are accepted are registered in the subscriber list, and each is sent a copy of the page. These nodes then behave like normal subscribers and commence participation in the subscriber and owner monitoring protocol.

If a node should fail during the progress of a transaction which it has originated, the database may be left in an inconsistent state. There are two aspects to be considered:

- a) Transaction recovery - dealt with in Section 2;
- b) resolution of structural inconsistencies.

The problems associated with (b) are resolved using the consistency checking protocol.

The consistency checking protocol runs continually as a background task on all nodes. It checks the structure of the database to ensure that it remains consistent. For example, for each index entry that it finds it will check that the relevant data record actually exists. Wherever an inconsistency is detected, it is corrected. In the above example, the index entry would be deleted if the data record could not be found.

1.4 USER INTERFACE

To the user the whole of the database appears to be resident on his own machine, thus the user is unaware of the distributed nature of the database.

The user interface comprises four access procedures:

- a) The Reliability Interface
- b) The Performance Interface
- c) Define Logical View
- d) Read Logical Record

The Reliability Interface

This interface provides the following features:

- a) Create a data record

This ADDAM call stores the data record in the appropriate free record slot in the appropriate page, and creates any necessary index entries.

- b) Update a data record

Note that update is at the unit of the segment. The update of a whole record is achieved as a series of separate updates to the segments in the record. Also a segment must be read, the images updated, then the update command issued, for the update to succeed.

- c) Delete a data record

Any index entries are deleted, along with the data record, when this function is executed so that consistency is maintained.

The Performance Interface

- d) Update a data record

This interface is used by application programs to cause a segment to be updated as rapidly, and with as little use of system resources as possible. There is no requirement to read the segment before updating it. There is no reply by ADDAM and thus no indication to the user of success or failure.

Define Logical View

This interface allows a user to define a subset view of the database. A user can specify a sequence in which a group of records should be retrieved and can attempt to make the retrieval or update as fast as possible by requesting local residency of data. The following commands are available:

- e) Define a logical view before reads

Passing qualifiers to the record name of interest gives two benefits. First of all, on subsequent reads by the application, only those records that meet the previously defined 'view' are returned to the application. Secondly, the local ADDAM will try to become a voluntary subscriber to the page of interest when the logical view is defined, so that read operations are faster. This can speed up the processing of both the get-next-record type of operation, and reliability updates which must be preceded by reads.

- f) Define a logical view before updates

This feature allows the application to request local copies of indexes to data records which are to be updated later.

Read Logical Record

The read function supports the reading of a single segment, up to 30 bytes, or a complete record currently up to 450 bytes in size.

g) Read data records sequentially

Used in conjunction with the definition of a logical view, this function supports the "get-next" operation. Unwanted records are filtered out by ADDAM, according to the 'view' selected.

h) Read a data record by key value

ADDAM supports a selective read, with the user specifying the record name and several qualifiers whose values allow ADDAM to retrieve only that record which fits the criteria.

2. MAINTENANCE OF CONSISTENT DATA2.1 THE RELIABILITY PROTOCOL

Early versions of ADDAM had a reliability protocol which, though ideal for the simplest of transactions, was always seen as inadequate for updates on complex data structures. The protocol avoids the requirement to lock records by forcing a user to read each segment before updating it; when two users wish to update the same segment, the first to present the update request succeeds and the other must read the segment again before updating. This protocol has now been redesigned to allow the user to define long transactions, that is multiple database operations that form a single commit unit. Although ADDAM does not support consistent readsets (i.e. reading a set of records as if from a consistent database frozen in time), the new protocol ensures that the duration of temporary inconsistency in the database is minimal and allows a transaction to be rolled back quickly if an error is detected.

For most reliability transactions, ADDAM recognises three phases: the distribution phase, the commit phase and the completion phase [Figure 3]. Where two transactions are running in parallel trying to update the same segment, the first transaction to reach the commit phase succeeds and forces the other to be aborted. These phases may most easily be explained by reference to an example: in the example chosen a user wishes to update a segment in each of two data records as a single commit unit.

Distribution Phase

Firstly the user calls ADDAM to define the start of the transaction - this causes ADDAM to abandon and roll back any transaction initiated by the user but not completed. ADDAM allocates a unique ID which it attaches to all messages fulfilling the transaction. The user calls ADDAM a second time presenting the update to the first record segment.

At the originating node, ADDAM determines from the schema whether the update to a data segment could have implications for other records (e.g. index records): it is possible that the segment may contain a secondary key, or may contain the partitioning field whose value determines on which page the data record is stored. In either case, ADDAM reads the live segment and compares the relevant fields in the live segment and the proposed update. Having determined which records (and which pages) are affected by the update, ADDAM decomposes the update into a series of fragments, each of which is an update request for one database segment. Each fragment is broadcast in turn to the node which owns the page in question.

On the owner page, a preliminary check is made to determine whether the fragment may proceed. If this check is successful, the owner node stores details of the fragment and broadcasts in turn to all nodes which subscribe to the page. Subscribers reply to the owner which in turn replies to the originator; if insufficient subscriber replies are received by the owner then this is reported to the originator which immediately abandons the transaction, broadcasting a "roll back" message to all nodes and replying to the user. On both owner and subscriber nodes, the fragment is stored in a temporary buffer until the transaction is either completed or rolled back. ADDAM replies to the user when all fragments have been distributed, and the distribution phase is repeated this time for the update to the second segment.

Commit Phase

When the user has presented all parts of his transaction for distribution, he calls ADDAM to commit the transaction to the database. In the commit phase, the originator sends a single broadcast message to all nodes. For each fragment, the owner node verifies that the requested change may be made successfully, and at this stage searches its list of transactions in the distribution phase to roll back any which will be caused to fail by the fragment being committed. Once the owner has committed the fragment, it broadcasts to the subscribers and counts their replies. The owner replies to the originator. The originator checks back replies for each page against a list of pages involved in the transaction. If owners fail to reply within a timeout period (e.g. if a new owner of a page is being elected), the originator broadcasts the commit message again to give "lost" owners a second chance to reply. If a successful reply has been received from all owners, then a final completion phase begins.

Completion Phase

In this phase the owner again broadcasts a single message to all nodes. For each fragment, each owner node makes live the updated segment, copying data from its temporary buffers to the data page. Each owner node instructs its subscribers to follow suit. At this stage the transaction is irrevocable but each participating node replies to the originator; the originator waits long enough to receive one reply before itself replying 'successful' to the user. If the originator does not receive a reply to its "complete" message, it retransmits the message. If still no reply is received then the outcome of the transaction is uncertain and the originator institutes consistency restoration (see Section 2.2).

If after some timeout period a node finds that a transaction is committed but not completed then it broadcasts an enquiry to other nodes "What happened to transaction XYZ?" Any node which has received either a "complete" message or a "roll back" message for the transaction replies to the uncertain node which then follows suit; if no reply is received then the uncertain node itself broadcasts a message to roll back the transaction and it rolls it back itself.

This protocol is expensive in terms of inter-node messages and processing within ADDAM; this expense is essential where the transaction affects more than one record but is unnecessary in the majority of cases where an update to a single segment is required. For reliability updates to a segment which does not contain a secondary key or a partitioning field, and which therefore can be confined to a single update fragment, ADDAM allows a user to commence, distribute and commit the transaction in a single call. Within ADDAM the distribute and commit phases are compressed. The originator broadcasts a "complete" message on receipt of a successful reply from the owner, just as in the three phase protocol.

The ADDAM reliability protocol gives an almost perfect guarantee that a transaction is consistently committed to the database or that the failure is reported to the user. The only circumstance in which the database can become inconsistent is when damage occurs to the network which breaks all highways or destroys nodes such that all copies of a page are lost; even in this remote circumstance ADDAM is capable of detecting the inconsistency and restoring the database to a self-consistent state, though this may result in some loss of data. The way in which this is achieved is described below.

2.2 RESTORATION OF CONSISTENCY

There are a number of situations in which it is reasonable to examine whether the database has become inconsistent. These situations are:

- a) When ADDAM attempts to read a data record using an index and finds that the index page does not exist;
- b) When ADDAM reads an index record and finds that the record pointed to by the index contains the wrong key value;
- c) When ADDAM creates a data record on an existing page but finds that no page exists for an index record to be created;
- d) When highway recombination occurs (see Section 3).

In each of these cases, ADDAM determines from the schema which page numbers need to be checked, and institutes for each suspect page a consistency restoration procedure. This involves:

- | | | |
|------------------------|---|---|
| for a data page | - | check that all indexes to each record exist; if any index is missing, create it; |
| for a prime index page | - | check that the data record exists corresponding to each index; if an index is found to be wrong, delete it. |

2.3 MANDATORY RELATIONSHIPS

If two data types A and B share a common attribute and are so closely related that a record of type B is incomplete without a matching record of type A, then there is said to be a mandatory relationship between A and B. We can distinguish two sorts of mandatory relationship: a 'loose' relationship in which an unattached type B record retains some significance in the absence of an associated type A record; and a 'tight' relationship in which an unattached type B record is nonsensical and indicates corruption or loss of part of the database. Appendix A gives some examples of mandatory relationships.

ADDAM allows the database designer to identify particular relationships as mandatory, provided that for any type B record there can exist only one matching type A record. Where a mandatory relationship has been defined, ADDAM prevents a type B record from being created unless the corresponding type A record can be found, and automatically deletes all corresponding type B records whenever a type A record is deleted, or is found to be missing during consistency checking.

3. SELF-REGENERATION AND NETWORK REPAIR

3.1 INTRODUCTION

As already described, ADDAM monitors the number of copies of each page and maintains enough copies to ensure survivability of data. Nodes holding subscriber copies of a page regularly report to the node which is owner of the page, and the acknowledgements sent by the owner are received by all subscribers so that each is aware of the existence of the others. If, through hardware failure, an owner node fails to respond to a subscriber, ADDAM chooses the most suitable subscriber to become owner in a process known as an election.

In a configuration with a single highway connecting the nodes, a break in the highway results in two part highways each with nodes running ADDAM software: two separate databases (consistent within themselves but not between themselves) are created on each side of the break, the missing owners being recreated by the election process and completely missing pages by the consistency restoration process. In a database manager like ADDAM it is in practice not possible to differentiate between the situation where the highway has broken and the situation where one or more nodes have been totally lost, because each node "sees" the system by use of the highway. Therefore the break in the highway means that users on each half will continue to alter data so that "duplicate" pages on each side become more and more dissimilar and inconsistent with each other. Each autonomous ADDAM database sees only some of the transactions current on the real-world entities represented in the database.

Although great importance is attached to consistency and survivability, of even greater importance is that the system should be available: in extreme circumstances some compromise on reliability is accepted in order to provide a continuing service to users.

Recent studies which we have carried out have considered what strategy should be adopted when a break in the highway is repaired and two different databases recombine. When a user updates a segment through ADDAM, the live segment is replaced in its entirety by the updated segment. This restriction enables ADDAM to avoid the consistency problems posed by concurrent updating. For this reason, ADDAM transactions are in general not commutative: there is no way that the transactions of two ADDAM databases could be resequenced on recombination to properly take account of the changes to the real-world entities.

On recombination, there will be pages for which there are two candidate surviving owners. For each page, one owner copy will continue as owner and the distinctive information held on the other copy will be deleted. In order to choose which owner must survive ADDAM has to consider two factors:

- a) Which owner is the most up-to-date or has had the most changes made to it.
- b) Which owner comes from the same half of the network as other related owners.

The second factor is more important even than the first because the self consistency of the surviving database relies upon it. There would not be much point in an owner copy of a data page from one half of the network surviving together with a related index page from the other half.

3.2 DATA GROUP NOTIFICATION

The problem is solved in the following way. Firstly, it is recognised that many record types and pages are mutually dependent and form a natural "data group". An example of a data group might be the following records:

- vehicle type data (mandatory)
 - one record for each type of vehicle giving general details
 - e.g. endurance, performance
- vehicle occurrence data
 - one record for each vehicle giving individual details
 - e.g. callsign, fuel state, weapon state
- vehicle occurrence data has
 - a prime index on key vehicle ID
 - a secondary index on key vehicle type

The database designer defines in the schema a data group for each record type and an independent record type for each data group - other types within a group are termed dependent. Secondly, each node regularly broadcasts to other nodes so that each is aware of the other nodes sharing the same part highway, and each is aware of any change in the nodes with which it is in contact. These broadcasts identify the nodes with which the sending node is in contact, and an "alteration factor" for each page local to the sending node. The alteration factor is the principal means of deciding between two candidate owner copies: it depends on both the number of updates made to the page and to the number of records stored. Thirdly the broadcast messages received at each node are sorted by data group so that the node can determine whether recombination has just occurred and whether any pages have multiple owners.

The following example shows how the data group monitoring protocol and the subscriber and owner monitoring protocol work together to maintain a consistent database. In the example, the network consists of three nodes X, Y and Z and recognises two data groups A (pages a1 (independent), a2) and B (pages b1 (independent), b2, b3). The following table shows the distribution of pages and shows which are owner and subscriber copies:

| | node | | |
|------|-------|-------|-------|
| | X | Y | Z |
| page | a1(s) | | a1(o) |
| | a2(s) | a2(o) | |
| | b1(s) | b1(o) | |
| | | b2(s) | b2(o) |
| | | | b3(s) |

Each node regularly broadcasts two messages, one for each data group. Each node receives six messages which it sorts according to data group. In case a break in the highway occurs, some quantity is required to identify the part-highway on which a node is located: from all the messages received a node selects the lowest node ID which acts as a crude part-highway ID and is included in each data group message that is broadcast. The data group messages contain the following information:

```

data group
node ID
lowest node ID
page number of page p
flag set if page p is owner copy
alteration factor of page p
page number of page q
flag set if page q is owner copy
alteration factor of page q
etc

```

The following problems may arise on the network:

Two Owners

Each node analyses the messages received, taking one group at a time. Initially when no break occurs in the highway, all nodes agree on the lowest node ID and it can be seen by inspection that each page has at most one owner. If a page has more than one owner than this is detected and quickly corrected. The owner with the greatest alteration factor is chosen to survive and the other owner is deleted. Node ID is used as a tie breaker so that a sole survivor may always be chosen.

A Break in the Highway

This situation is detected as part of the subscriber and owner monitoring protocol (described in Section 1.3). If a break were to separate X and Y in one half from Z in the other, then the following changes could occur. Firstly, lone subscriber pages elect themselves owners, and new subscribers are coerced on other nodes. The new page distribution is:

| | node | | |
|------|-------|-------|-------|
| | X | Y | Z |
| page | a1(o) | a1(s) | a1(o) |
| | a2(s) | a2(o) | |
| | b1(s) | b1(o) | |
| | | | b2(o) |
| | | | b3(o) |

The database on X and Y in fact survives intact. Initially Z is owner of the independent page a1 and the dependent pages b2 and b3. Consistency restoration then decides that there is no value in preserving b2 and b3 without the independent page b1, and so b2 and b3 are deleted. An empty page a2 is created when an application on Z needs to create a new record.

Highway Recombination

Network repair is detected when a data group message received contains a lowest node ID different to that recognised by the receiving node. Again, each data group is considered separately. We believe that the alterations made to the data group are more significant than the updates to a single page. The problem in recovering from network repair is essentially that of deciding, for each data group, which part-highway's pages are to be kept. To preserve self consistency on recombination, it is essential that the set of all pages in a data group should be chosen from the same part-highway: the example in Appendix B shows why this is the case. Having identified the "best" part-highway, pages (of the given data group) from other part-highways must be deleted because in general there is no way of detecting or resolving anomalies between pages from different part-highways.

If a recombination of the network is detected in which all pages of the independent data type within a group occur on one part-highway, then pages from that part-highway are chosen to survive. If pages of the independent data type within a group occur on more than one part-highway, then the average alteration factor for data pages is calculated for each part-highway, each page number being counted once only if more than one copy is found: the pages from the part-highway with the greatest alteration factor are chosen to survive. If the alteration factors for pages in our example are as follows:

| | node | | |
|------|-------|-------------------------|----------------|
| | X | Y | Z |
| page | a1(4) | a1(4) b1(3) b2(5) | a1(2) a2(8) |
| | b3(7) | | |

then the alteration factor (data group A) for part XY is 4, the alteration factor for part Z is 5, and the copies of a1 and a2 on node Z will survive.

In the following further example, the network consisting of P, Q, R, S and I is formed by joining P, Q and R with S and I. Data group C consists of page numbers c1 (independent), c2 (independent), c3 and c4. The table below shows the distribution of pages together with their alteration factors:

| | node | | | | |
|------|----------------|----------------|-------|----------------|-----------------|
| | P | Q | R | S | I |
| page | c1(1) c2(4) | c1(2) c3(9) | c2(4) | c1(5) c4(9) | c3(10) c4(8) |

If the highest alteration factor is counted for a page that occurs more than once on a part-highway (owner and subscribers may not be synchronised), then the alteration factor (data group C) for part PQR is $(2 + 4 + 9)/3 = 5$ and for part SI is $(5 + 10 + 9)/4 = 6$. This time, copies of c1, c3 and c4 on nodes S and I will survive.

We conclude that when network repair takes place, some loss of data is inevitable and that the strategy described above minimises the consequences of that loss.

4. CONCLUSION

In high resilience applications we believe that a reliable distributed database management system is vital but it brings with it significant costs in terms of computer and network usage. Our detailed designs show that this cost is not prohibitive, and we believe that the pressure for greater resilience and system modularity will demand the adoption of distributed data management techniques throughout military real-time networks.

REFERENCES

- [1] DEFENCE STANDARD 00-19 THE ASWE SERIAL HIGHWAY, Ministry of Defence, 1981
- [2] TILLMAN, P. R. ADDAM - The ASWE Distributed Database Management System, Proceedings of the Second International Symposium on Distributed Databases, North-Holland Publishing Co., 1982.

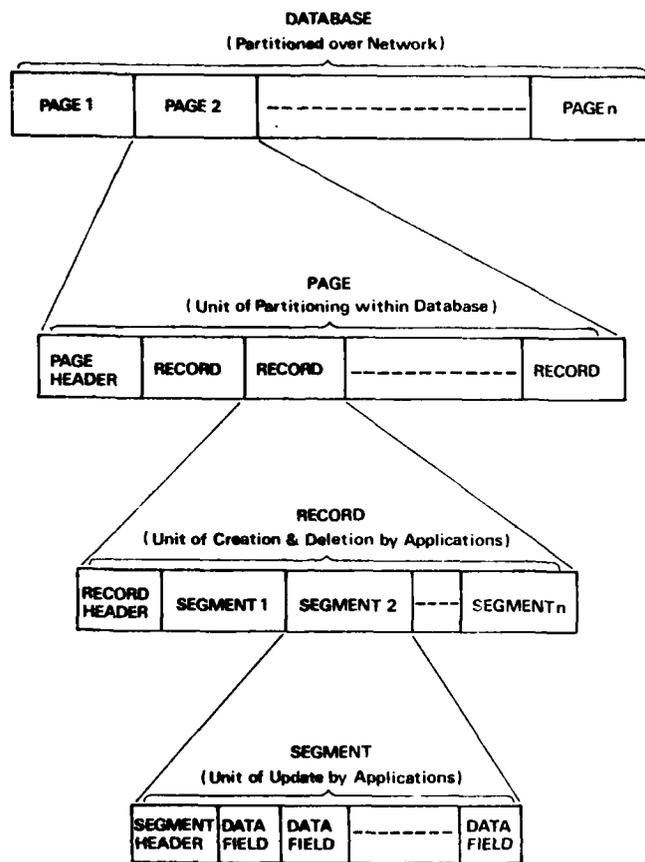


Figure 1 - SUBDIVISIONS OF THE DATABASE

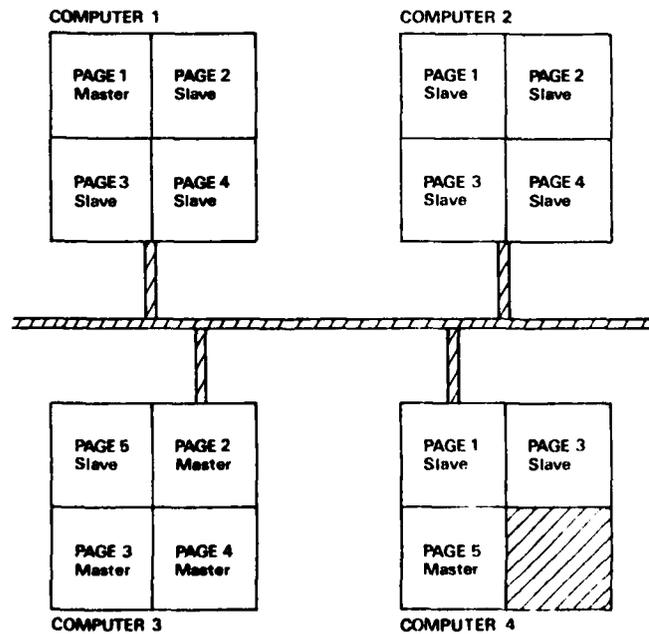


Figure 2 - PARTITIONING AND REPLICATION

DISTRIBUTION PHASE

Originator broadcasts individual fragments to owners
Each owner broadcasts to its subscribers
Subscribers reply to their owner
Owners and subscribers store the fragments
Owners reply to originator

COMMIT PHASE

Originator broadcasts to all nodes
Owner nodes roll back any conflicting transaction
Each owner broadcasts to its subscribers
Subscribers reply to their owner
Owners reply to originator

COMPLETION PHASE

Originator broadcasts to all nodes
Owner nodes copy fragments into database
Each owner tells its subscribers to do the same
Subscribers reply to their owner
 (owner need not wait for them)
Owners reply to originator
 (originator need only wait for one reply)

Figure 3 - THE RELIABILITY PROTOCOL

APPENDIX A Examples of Mandatory RelationshipsExample 1: loose

type A - track type:
one record describing attributes of objects a radar can detect
(e.g. DC-10, Exocet missile).

type B - radar track:
one record describing details of object detected by the radar
(i.e. position, height, speed and type).

Clearly every radar track should have an associated track type (even if only "unknown"). However if the track type data is lost (due to a highway break for example), the radar track data is still significant.

Example 2: tight

type A - job description:
one record describing the details of a task being carried out by members of a ship's crew.

type B - crew member - job:
one record associating a crew member with a task.

Clearly it is meaningless for a crew member to be doing a job which has no description (type A being assumed to cover such items as "miscellaneous", "sick" etc), and the record "crew member - job" is of no value if the "job description" data is unavailable.

APPENDIX B Examples of Highway Recombination

type A - track data:
one record for each object detected by a sensor.

type B - track-track pairing data:
one record for each "pairing" relationship between two tracks.

There is a mandatory relationship between record types A and B - a pairing record is just nonsense if either of the paired tracks does not exist.

Suppose that page a1 contains records of type A and page a2 contains records of type B. A break occurs in the highway so that each part-highway has its own copy of pages a1 and a2.

In each part-highway, new track records are created. After a time, track records with the same ID in the two part-highways point to different objects. In one part-highway track 99 is paired with track 1; in the other part-highway track 99 is paired with track 2.

When recombination occurs, the wrong tracks will appear to be paired unless the surviving owner copies of pages a1 and a2 are chosen from the same part-highway.

**SIMULATION - A TOOL FOR COST-EFFECTIVE SYSTEMS
DESIGN AND LIVE TEST REDUCTION**

by

Roland Gauggel, dipl. Phys.,
Missile Division
Bodenseewerk Gerätetechnik GmbH
Postfach 11 20
D-7770 Überlingen Bodensee
West Germany

Taking advanced passive infrared guided missiles as an example the author claims in his paper that missile system simulation - both software and realtime hardware-in-the-loop including background - is a valuable tool to find cost-effective system designs and also to drastically reduce costs of field testing and live firing trials. The author comes to the conclusion that the development of complex missile systems becomes questionable from a cost standpoint if the majority of the increased test efforts for this type of missiles is not substituted by missile system simulation.

The author addresses Bodenseewerk's missile system simulation philosophy, simulation methods, high level programming language and the interfaces between the involved hardware and software. An in-depth discussion of the influence of simulation onto the flight testing requirements of missile developments and the resultant cost savings conclude this paper.

My contribution to the subject of 'Cost-effective Guidance and Control Systems' pertains primarily to missile system simulation as a tool for cost-effective system design and reduction of extensive and expensive live test campaigns and not to a low cost realization or production of a specific design.

There is a great number of different domains of application for guidance and control systems implying many different requirements. I would like to concentrate in this paper on the field of unmanned weapon systems and here especially on tactical, passive IR homing missiles. The guidance and control problems occurring in such missiles are very complex and the extreme requirements to be met stress the limits of physical as well as technical feasibility. Some of these requirements such as

- fire and forget
- multiple target discrimination
- low miss distance
- large range

require autonomous guidance, at least in the terminal flight phase.

Presently available homing heads are governed by certain physical and technical limitations. The required mission ranges, for instance, are much larger than today's homing head acquisition range performance above all in adverse background conditions. This forces us to take a new approach, e.g. by introducing

- new homing heads with improved performance that implies upgraded intelligence
- and
- an inertial midcourse guidance phase.

The introduction of an inertial midcourse guidance, where the missile receives certain pre-launch target information from the aircraft and navigates inertially until seeker lock-on, circumvents homing head acquisition range limitations. (Fig. 1) For long missile flight times and/or extreme dynamic engagement conditions an updating of the target information would be necessary.

It is obvious that the introduction of midcourse guidance drives the complexity and therefore the costs up instead of down. There is also very little potential to offset this cost increase by re-arranging sensors and shifting some of their tasks into the computer software. Fig. 2 shows an example.

In the conventional approach the seeker head and inertial reference system are basically two separate units. The inertial package is used only until the seeker head has acquired the target; then it serves as a sensor package for the autopilot function. The seeker head itself is designed as a gyro-stabilized platform thus representing an inertial reference system. In today's possible strapdown approach the midcourse and the seeker head stabilization are done with a single inertial reference system using heavy fixed sensors. Thus, the inertial reference system becomes part of the seeker control and retains its full function throughout the entire flight. The line-of-sight rate information required for guidance, which in the conventional system is provided by the seeker gyros, is calculated for this implementation.

This solution allows a slight reduction of production costs as a result of the reduced number of sensors and other mechanical equipment. A considerable cost reduction can be realized from these measures, since other components, such as angle pick-offs have to meet much tighter specifications, resulting in component price increases.

All this shows that such measures have only a minor effect on cost and this effect is even less significant when it is considered that the costs for the inertial reference system are relatively low compared to the increasing homing head costs. For inertial mid-course guidance a minimum number of rate gyros/free gyros and accelerometers are always required, independent of where they are located.

The requirements for

- maximum target acquisition range in adverse background and IR countermeasure conditions
- multiple target discrimination

lead to increased complexity and costs in the field of homing head signal processing, particularly as the homing head must be able to perform automatic lock-on in flight due to its limited target acquisition range. Homing heads of the present generation which are mostly equipped with mechanical reticles do not meet these requirements. Imaging homing heads, however, promise the potential of providing a solution. Image processing implies the necessity to use complex and costly signal processing electronics combined with enormous computing capability. Here we face the direct conflict between the requirement of improved performance on the one hand and cost reduction on the other hand. The development of these advanced missile systems requires extreme technical and testing efforts in order to achieve the development aim at all. Particularly the testing effort will be an order of magnitude higher than with present generation IR missiles.

In my opinion the use of advanced simulation methods and facilities relieves the situation. This not only because simulation enables the successful development of such missile systems at all, but also because it actually provides the adequate means for design optimization, for keeping the development time within reasonable limits and for drastically reducing the expensive test phases which are otherwise necessary during development.

As it was discussed in detail during the AGARD National Day September 1983 in Munich, two types of missile system simulation are important (Fig. 3)

1. All math models simulations
2. Hardware-in-the-loop (HWIL) simulations.

In today's modern missile development the all math model simulation procedure uses a purely digital computational approach. In digital simulations the equations are numerically integrated requiring extremely fast computing capability. These extremely high computing speeds are realized by using not only one large digital computer but an entire computer network consisting of one general-purpose computer and several special-purpose processors attached to it. This multi-processor system allows for the partitioning of an overall simulation program into separate parts where each of these complex and time critical sub-programs is assigned to one of the high-speed satellite processors. This approach provides engineers the chance to include all essential guidance loop components, missile control and target/background representation, programmed on a physical basis including detailed error effects.

HWIL simulation offers the advantage that the hardware replaces the model equations allowing to test individual subsystems under closed loop conditions so that the math models can be validated under realistic operational conditions. In all those cases where the exact mathematical description of the physical system is difficult or even impossible the HWIL simulation is the only method. The hardware is represented by two different subsystem categories: analog equipment and digital computers. These hardware components are coupled to the multiprocessor system and/or general purpose computer by interface computers.

The simulation tasks to be performed during missile development are listed in Fig. 4. The successful execution of these tasks requires a number of all-math simulation models as well as HWIL programs of different levels and degrees of complexity which must be adapted to the specific problems. The functional test has to be performed under as realistic conditions as are possible in the laboratory. For this purpose the different subsystems are mounted on a flight motion simulator which performs, in realtime, all rotational accelerations and velocities of the missile according to the operational conditions.

When considering non-decoyed targets, and excluding complex target/background situations, the comparison with live firings clearly shows that the all-math and HWIL simulation provides an accurate description of the measured missile behaviour. The measured values differ from the simulated values by less than 10%; this applies also for missiles having a sensitive flight behaviour and operating under extreme initial conditions. The critical problem is to provide exact target representations including the multitude of natural background scenarios and artificial decoys.

The target and background representation has, however, been somewhat neglected in the past. Its importance has never been underestimated but the target characteristics, natural background and artificial decoys can vary to such an extent that a realistic representation of all possible combinations is almost not feasible. In the field of passive IR guided missiles Bodenseewerk introduced, some time ago, a new method which allows a much more realistic, realtime simulation of target/background, also under dynamic conditions. I presented this during the AGARD National Day September 1983 in Munich.

When using this high-performance target/background simulator the simplified target and seeker head models shown in Fig. 5, which are mainly based on simple transfer functions and approximate functional correlations, can be replaced by detailed models. Since the targets and the background are represented as a physical model, the signal processing can be described in a physical manner as well. As a result the approximation functions are replaced by the simulation of the actual process. The example to the right shows a mechanically scanning seeker model using a linear array detector.

Only a simulation technique of this type provides the capability to define the seeker target acquisition characteristics against complex backgrounds and to generate the actual seeker head error and/or guidance signals over the entire flight phase in the laboratory. It also provides answers to our questions concerning seeker head performance when acquiring decoyed targets and estimates of the accuracy of the aim point selection during the terminal guidance phase.

The seeker output signals are fed into the 6-DOF simulation model and used in both the all-math and the HWIL simulations in order to answer the all decisive question about miss distance (Fig. 6). In order to perform HWIL simulations both the target/background simulator as well as the seeker head signal processing must be run in realtime. This simulation technique is the only method to develop, test and optimize the signal processing algorithms in the laboratory. The result is a money saving effect since a high number of expensive flight trials are no longer necessary and development time can be reduced considerably.

A further important factor for cutting down on development cost is the compatibility and easy handling of software and hardware during simulation. The entire simulation complex basically consists of a conglomerate of

- simulation computers (general purpose, process control computers)
- analog and digital missile assemblies
- peripheral test and supply equipment (flight motion simulator, target simulator, auxiliary equipment)
- software/hardware interfaces

A vital element is the simulation language. Experts agree that the use of a high level simulation language has a distinct time and therefore money saving effect when this language is tailored to the specific missile design and simulation problems. Since there is no such specifically tailored language available on the market Bodenseewerk has developed its own language which is called BOSIM 80 and which drastically improves and accelerates our simulation.

The advantages of high level languages for the numerical integration of differential equation systems by using problem-oriented instructions, as well as the advantages of easy model equation notations and of comfortable input/output and sorting routines are generally accepted. Again, details have been discussed during the above mentioned AGARD meeting in Munich.

The availability of language instructions for easy handling of missile hardware and periphery is the decisive factor.

Let me now say a few words about Bodenseewerk's missile simulation philosophy, a loosely-coupled multiple processor system with its associated peripheral equipment tailored to the particular needs of missile system simulation - both software and HWIL. Missile hardware, flight simulators and support equipment are connected to a computer network by a software-controlled crossbar switch bus system, allowing any combination of processors, peripherals and HWIL equipment. The underlying philosophy for this type of network is that the required high throughput timesharing environment for simulation program development and the high speed single task environment for HWIL execution could hardly be realized on a single computer but rather would require at least three different computers. (For program development, for realtime program speed optimization, for realtime HWIL execution). Satisfying these requirements by installing separate computers would cause serious problems in data exchange, software maintenance and software configuration control to ensure software consistency throughout the missile division, as all these separate computers would have their own data bases, their own operation systems, languages and libraries. (Fig. 7, left hand side).

A better solution is to couple special computers of the same basic type using the same operating systems, identical languages and accessing a common data bank that holds all required programs, macros, libraries and data sets. (Fig. 7, right hand side). A benefit of that type of network is that there no longer exists a potentially error prone information flow between the computers.

In order to match the specific requirements of the different stages of simulation, beginning with program development and ending with the HWIL experiment, the computers must have special different capabilities. This is accomplished by attaching slave processors, thus adding the requested capabilities for the different environments. Fig. 8 shows the functional schematic of such a network, consisting of network processors of the same type, attached slave processors and special peripherals for added capabilities, and network for terminal and disc support.

Seen from the network's point of view, all network processors still are of the same type, allowing easy network management. Seen from the users' point of view, there are several available programming environments with the requested specific capabilities, as all users can access all processors via the terminal network, all processors can access all discs via the disc network and all users can demand for specific slave processors and special peripheral equipment via the software controlled crossbar switch system.

Let me now turn to the final issue of my paper, namely to the influence of simulation on missile testing (Fig. 9). The development of a missile system is always performed to a development specification which in turn is based on tactical and operational requirements. As soon as hardware of complete missiles becomes available actual firings will be started in order to evaluate the missile's performance against the development specification. Several different test phases follow each other, from contractor tests to operational tests and evaluation done by the official services. The actual performance and qualification proof against the development specification by missile firings is one of the key activities during development in order to demonstrate the successful completion of development to the customer.

The problem here lies with the complexity and diversification of scenarios in which the missile would have to be tested in order to demonstrate its overall mission performance. In this context not only the high number of necessary firings and the related high costs for the missile hardware are to be considered but also the technical and financial efforts which are required for presenting the targets in many different tactical scenarios. In particular, the consideration of all the possible different background conditions as well as artificial decoying and different attack formations drive costs to a prohibitive level.

Therefore, it becomes obvious that the test programs have to be tailored in such a way that only a limited number of selected conditions are proven by actual firings. The performance proof using missile hardware will thus refer only to a few samples out of the almost unlimited non-linear operational space. For cost reasons the number of selected samples can never be as high as necessary in order to provide an adequate performance proof for the customer. Here simulation is an adequate and by the way the only tool to improve the situation.

Parallel to the missile development a simulation model is set up which has been validated by extensive HWIL tests of all major subsystems. At the beginning of the test phases validated all-math and HWIL simulation programs should be available describing the complete missile. As already detailed before a target/background simulator is required in addition to the missile simulation programs.

When testing the complete missile the warhead is in most cases replaced by a telemetry package which transmits the main functional and operational data of the missile during the mission. These telemetry data as well as the known launch and target/background conditions are used for validating the related all-math and HWIL simulation results. When the main missile parameters, functions and signal flows comply with the results of the live firings within a certain tolerance range the missile simulation model can be considered to be validated and can thus be used for extensive performance calculations for the different target/background configurations. The actual performance proof of the development specifications and the operational requirements is then done by all-math simulations in the laboratory. This is a much more cost-effective means of performance verification.

To summarize, the advanced missile system simulation represents a mighty tool for money savings during the course of complex missile development. As discussed it provides adequate means for design optimization and reduces drastically the live test efforts. As far as production is concerned the savings effect is with the mechanization of an optimal design which leads to a minimum of hardware and test time with respect to development specification compliance.

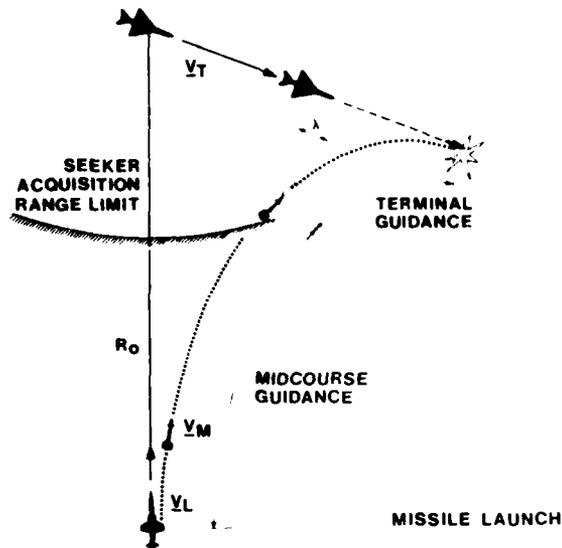


Figure 1 Midcourse guidance

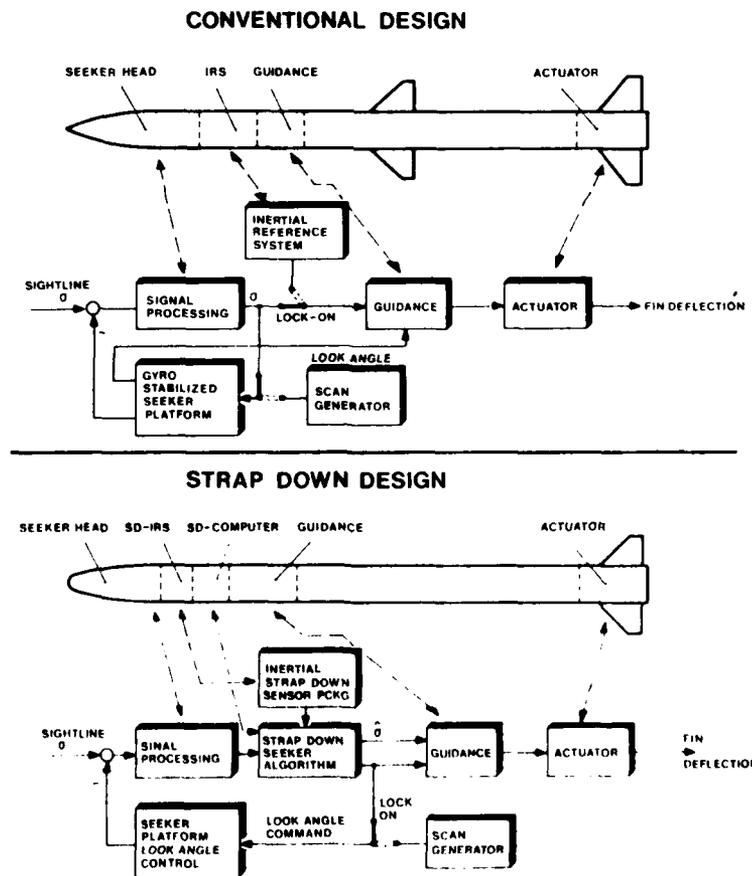
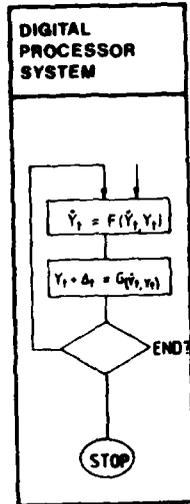


Figure 2 Realization midcourse guidance

ALL MATH MODEL SIMULATION



HWIL-SIMULATION

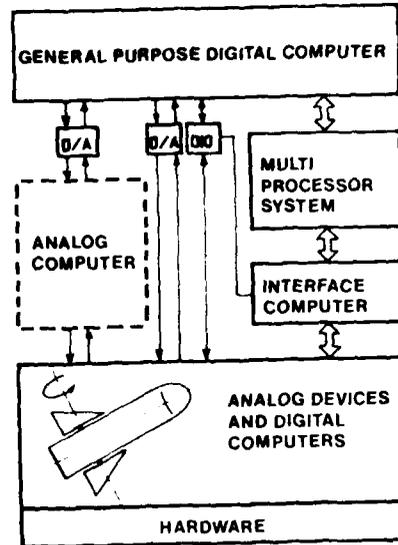
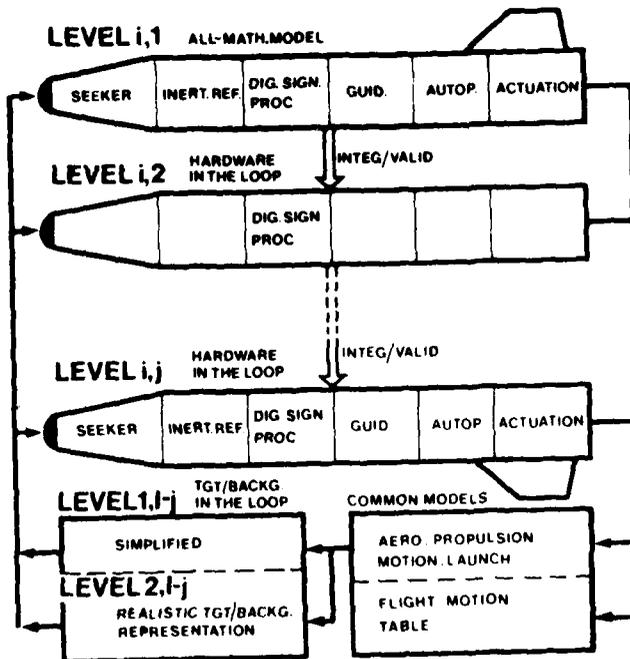


Figure 3 Advanced simulation techniques



FROM DESIGN TO SYSTEM EVALUATION & QUALIFICATION

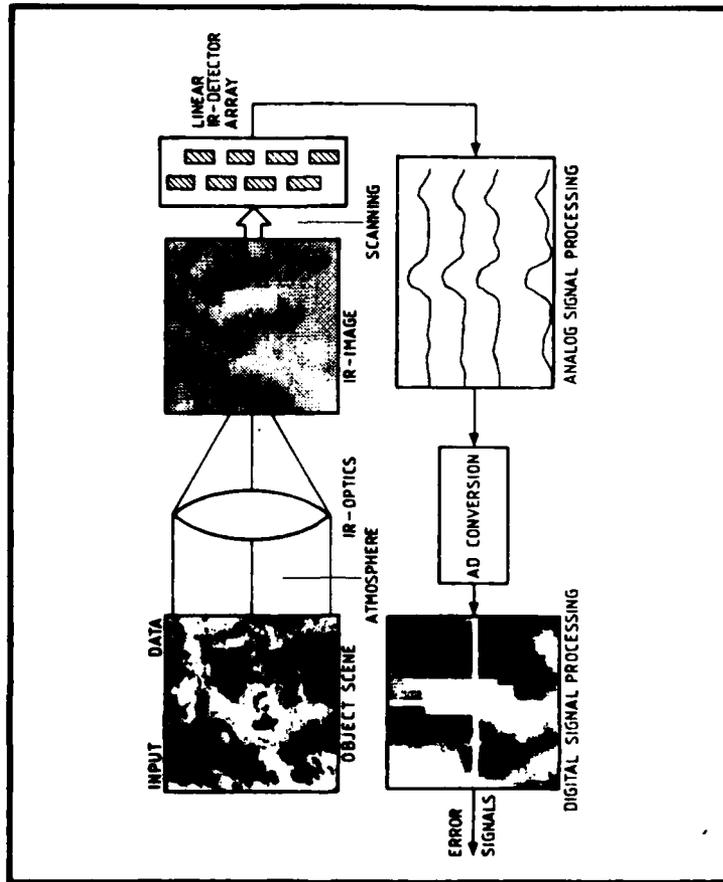
VARYING COMPLEXITY FOR DIFFERENT SIM PURPOSES

MAJOR SIM TASKS:

| | | |
|-----------------------------------|----|----|
| performance analysis/optimization | .. | .. |
| flight dynamics | .. | .. |
| aimpoint | .. | .. |
| param sensivity | .. | .. |
| miss distance | .. | .. |
| tolerance | .. | .. |
| range safety | .. | .. |
| preflight | .. | .. |
| postflight | .. | .. |

Figure 4

DETAILED IR SEEKER MODEL



SIMPLIFIED IR SEEKER MODEL

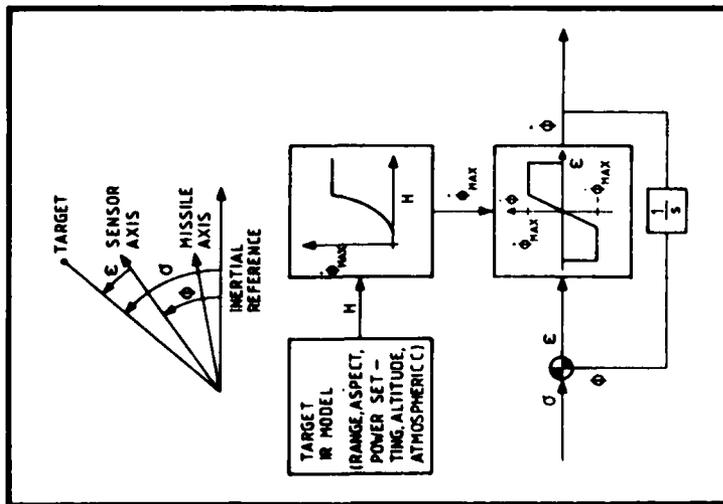


Figure 5

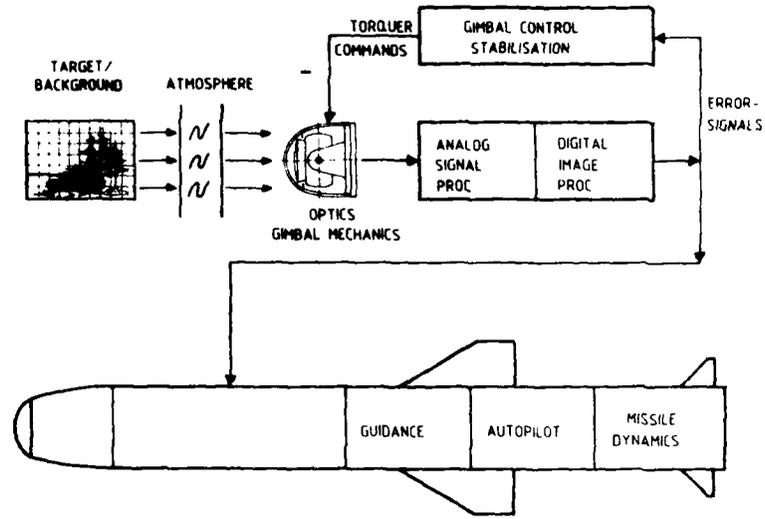


Figure 6 Full scale simulation model

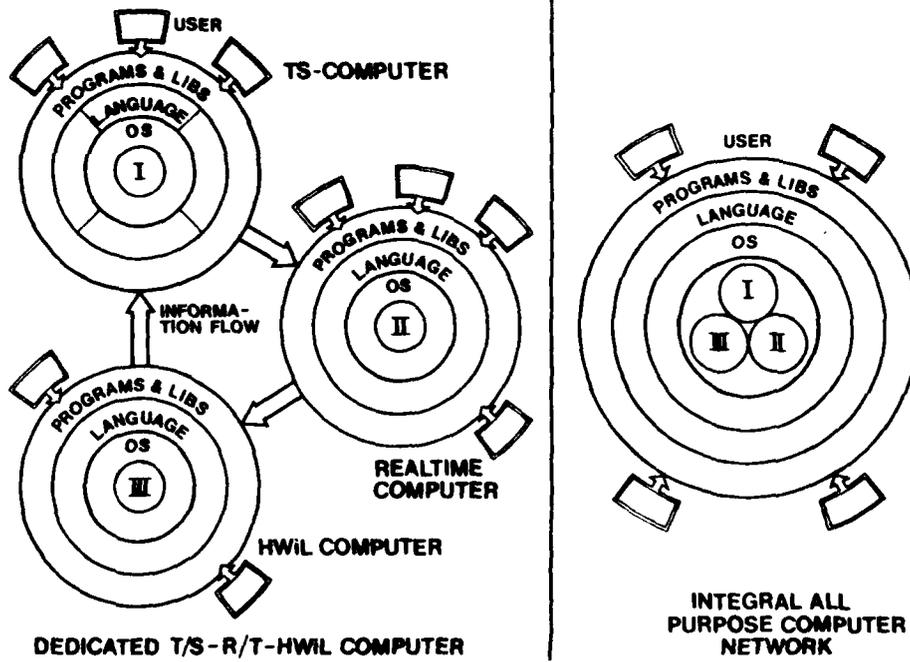


Figure 7 Data flow in different computer networks

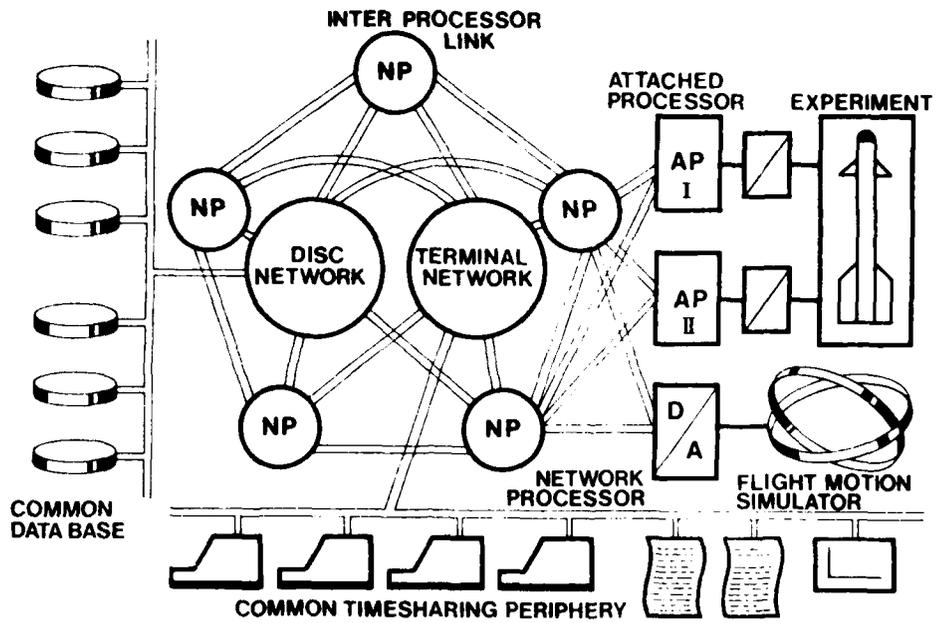


Figure 8 Homogenous network architecture

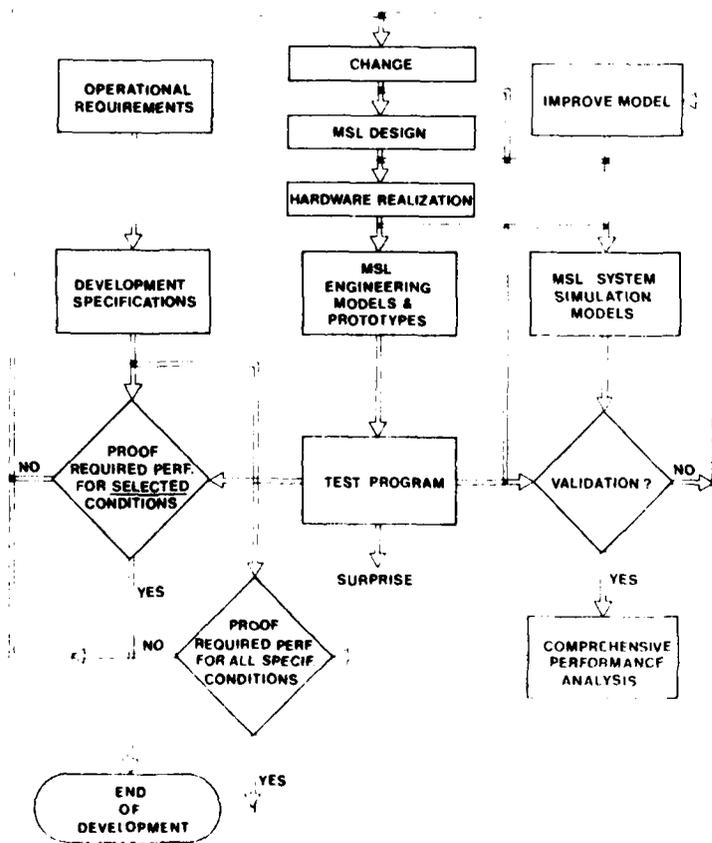


Figure 9

DESCRIPTION DU PROCESSUS D'ELABORATION DU DIALOGUE"HOMME MACHINE" DANS LES AVIONS D'ARMES EQUIPES DE TUBES CATHODIQUESROLES ET DESCRIPTION DES MOYENS DE SIMULATION PILOTEE

Monsieur Pierre HELIE
 AVIONS MARCEL DASSAULT - BREGUET AVIATION
 78, Quai carnot 92214 SAINT CLOUD

FRANCE

RESUME

La conception du poste d'équipage des avions d'armes actuels et futurs est soumise à différents types de contraintes qu'il est possible de résumer comme suit :

- Objectifs de réalisation (charge de travail, fiabilité, sécurité ...)
- Contraintes de réalisation (cahier des charges, avionnage, ...)
- Etat de la technologie de commandes et visualisations

L'analyse de ces contraintes et les développements technologiques en cours autorisent à prédire sans trop de risques que les avions d'armes du futur proche auront une planche de bord équipée intégralement de tubes cathodiques (ou de systèmes permettant des présentations équivalentes), associés à quelques instruments conventionnels pour l'ultime secours, et un système de commandes constitué entre autres de commandes à labels variables réparties sur les écrans, de commandes physiques multiplexées implantées sur Manche et Manette et d'un système de dialogue vocal.

Au niveau des ingénieurs chargés de l'étude du dialogue "Homme-Machine" l'apparition de ces nouvelles techniques a provoqué la nécessité d'une part de remettre en cause les principes établis (par exemple suppression des instruments conventionnels, utilisation du HUD comme instrument principal de pilotage) et d'autre part de développer de nouvelles méthodes de travail car il n'est plus possible de procéder par petits pas à partir de la réalisation précédente (multiplexage et juxtaposition des informations, optimisation des présentations aux phases de la mission) chaque poste étant optimisé pour un avion donné.

Les AMD.BA, pour répondre à ces besoins, au vu de leur expérience sur les avions actuels (militaires et civils) et dans le cadre d'études de postes d'équipage futurs ont mis au point un canevas de tâches et documents à réaliser permettant aux responsables d'identifier les différentes étapes du développement et d'effectuer quand il est encore temps les rebouclages entre les différentes parties en présence (bureau d'étude, équipementier, utilisateur).

L'ensemble des documents et tâches à réaliser peuvent se décomposer en :

- documents de définition
- documents liés à la réalisation
- tâches de validation liées à la définition et à la réalisation.

En ce qui concerne l'élaboration du dialogue "Homme-Machine" le présent exposé ne s'intéresse qu'aux documents et aux tâches liés à la définition.

La suite de la présentation se propose de détailler chacun des documents à réaliser (objectif, contenu et moyens) et de présenter pour les tâches de validation, les moyens mis en oeuvre pour les réaliser ainsi que les différentes utilisations possibles de ces moyens.

DOCUMENTS A REALISER

Document "Modes et Commandes"

Objectif du document

L'objectif principal de ce document dont la réalisation suit immédiatement la phase avant projet est de fournir à l'utilisateur de l'avion (client, essais en vol) une enveloppe des possibilités du "système avionique" de l'avion. En retombée il doit permettre de fixer au concepteur un cadre de travail pour l'élaboration du dialogue homme-machine.

Ce document est le fruit d'un travail collectif auquel participent les différents spécialistes concernés (Matériel, Fonctionnel, Utilisateur).

Contenu

Le "Modes et Commandes" comporte deux parties orientées "Matériel" d'une part et "Fonctionnel" d'autre part.

La partie "Matériel" présente tous les équipements ou systèmes "fonctionnels" du système avionique ; ces systèmes "fonctionnels" sont essentiellement des capteurs qui peuvent posséder divers modes de fonctionnement chacun de ceux-ci pouvant être utilisés par une fonction du système avionique ; ce seront par exemple le radar, la ou les centrales à inertie, les capteurs optroniques ...

Pour chacun de ces systèmes le document présente une description physique de l'équipement, de ses modes de fonctionnement et de ses commandes spécifiques.

Les équipements d'interface (commandes et visualisations) font également l'objet d'une description physique permettant d'évaluer leurs capacités globales (par ex. modes de balayages, nombre de couleurs, champs pour les visualisations).

La partie "Fonctionnel" du document se décompose elle-même en deux sous-ensembles :

- le premier de ces sous-ensembles présente une philosophie générale d'utilisation de la cabine par l'intermédiaire des systèmes de commandes et de visualisations, il comprend une description générale applicable dans tous les modes des principes de dialogue homme-machine.

Il décrit :

- . les principes d'affectation des commandes "manches et manettes"
 - . les principes de dialogue à partir des différents systèmes disponibles (commande vocale, clavier multi-fonction, écrans) ; pour les écrans il présente des règles générales d'utilisation (définition de zones, affectation de ces zones, ...)
 - . des descriptions présentant la philosophie de présentation des panneaux et alarmes.
- Le second sous-ensemble est une présentation de haut niveau des différents modes de fonctionnement du système avionique.

Chacun des modes de fonctionnement du système avionique est décrit par :

- le moyen d'appel du mode et de ses éventuels sous-modes
- les principales commandes et actions possibles
- l'affectation des visualisations comprenant pour chacune d'elle le type et l'énumération des informations présentées
- les moyens de sortie du mode

Enfin ce second sous-ensemble dresse sous forme de tableau ou chronogramme un récapitulatif des différents modes, de leur priorités respectives, des compatibilités, et des différentes transitions possibles entre modes et commandes correspondantes.

Moyens

Ce sont essentiellement (outre le crayon et la gomme) des moyens humains qui sont mis en oeuvre pour la rédaction de ce document.

Document de "Spécifications Globales"

Chacun des modes ou des sous-modes de fonctionnement mis en évidence par le document "Modes et Commandes" fait l'objet d'un document de "Spécifications Globales".

Objectif

L'objectif de chacun de ces documents est à partir du cadre de travail fourni par le document Modes et Commandes de faire une description dans le détail des commandes et visualisations associées aux différentes fonctions opérationnelles activées dans ces modes. Le document est réalisé par des spécialistes fonctions.

Bien entendu compte tenu de l'activation simultanée de plusieurs fonctions opérationnelles un "juge de paix" système arbitre les éventuels conflits et priorité.

Contenu

Ce document présente dans le cadre d'un fonctionnement normal du système (procédures conventionnelles, bon fonctionnement des équipements nécessaires à la fonction) une description précise et exhaustive de l'ensemble des actions équipages et des visualisations associées à la fonction en objet.

Ce document est en fait, en finale, le "Manuel Pilote" de la fonction décrite.

Moyens

Compte tenu du caractère du document de Spécification Globale, de l'utilisation généralisée des tubes cathodiques et de la forme répétitive de certaines visualisations, il a été ressenti un besoin d'aide au concepteur basé sur un système informatique et graphique.

Le matériel utilisé aux AMD.BA pour ce besoin est un système actuellement en place pour répondre à d'autres besoins (synoptiques systèmes, câblages électriques, éditions de documents). Les principaux avantages de ce système sont :

- ses capacités intrinsèques
 - menus de symboles
 - appel ou suppression des fonds graphiques
 - choix d'échelle
 - "hard copy"
- le travail sur écran cathodique (monochrome et éventuellement couleur) de façon à placer le concepteur dans un environnement "représentatif" des tubes présents sur l'avion
- archivage et édition des visualisations

TACHES DE VALIDATION - SIMULATIONS PILOTEES

Le paragraphe précédent a présenté les documents à réaliser pour définir l'utilisation de l'avion associé à son système avionique, bien que n'étant pas rentré dans le détail de chacune des fonctions il ne fait aucun doute sur la complexité de ces fonctions et sur la difficulté de réaliser la validation des spécifications à partir d'un seul document aussi bien fait soit-il.

Il convient également de prendre en compte à ce niveau des contraintes de temps et d'argent pour arriver à la conclusion sur la nécessité d'effectuer juste avant la réalisation (c'est-à-dire rédaction des spécifications détaillées et réalisation du logiciel) une validation de ces spécifications pour, autant que possible, limiter les essais au sol ou en vol avec les matériels réels à des contrôles de la réalisation et non pas des spécifications ; c'est tout naturellement que les AMD.BA ont été conduits à la mise en oeuvre de simulation de systèmes en temps réels permettant d'une part d'étudier l'aspect commandes et visualisations et d'autre part de valider certains algorithmes complexes de guidage ou autres. Ce système OASIS opérationnel actuellement aux AMD.BA a été développé pour le Mirage 2000 et est adapté pour tous les nouveaux avions en développement (militaires et civils).

Configuration du système

Le système OASIS a été conçu avec le double objectif, d'une grande souplesse d'utilisation permettant de s'adapter à des travaux de type étude (qui exigent facilité et rapidité dans la réalisation de modifications) et d'être économique tant dans sa réalisation que dans son exploitation.

Ce double objectif a conduit à concevoir un système réalisant une simulation complète des équipements et du système. De même les interfaces physiques homme-machine sont réalisées avec cet objectif et ne sont donc pas nécessairement les interfaces de l'avion étudié. Compte tenu de la tendance à la disparition des postes de commandes spécifiques les soucis de représentativité des interfaces devrait s'effacer dans le proche avenir.

La simulation est réalisée à partir de 3 ensembles comprenant :

- un ordinateur avec ses périphériques qui reçoit les informations provenant des actions du pilote effectue les calculs de simulation de l'avion au SNA et gère le système graphique
- un système graphique qui à partir des messages provenant de l'ordinateur trace la symbologie sur un tube cathodique
- une console pilote qui supporte le système graphique et regroupe les différentes commandes de mise en oeuvre du système y compris celles présentées sur le manche et la manette des gaz.

Description du système OASIS

Le système OASIS est d'une part composé d'un ensemble matériel et d'autre part d'un ensemble de logiciels spécifiques développés pour les besoins de cette simulation.

Le matériel

Le calculateur et ses périphériques

Le calculateur utilisé est un ordinateur 32 bits bien adapté pour les opérations temps réel. C'est essentiellement cette caractéristique, associée à l'expérience d'utilisation de cette machine pour le traitement des données d'essais en vol, qui a guidé le choix vers ce système.

Il possède deux unités centrale CPU et IPU dont une, le CPU, est affectée au traitement des entrées/sorties. Cette architecture permet dans certaines applications gourmandes en temps de calcul de dérouler des morceaux de programme en parallèle dans chacune des unités centrales et donc de gagner en temps de cycle.

Le calculateur possède les périphériques suivants :

- l'ensemble de visualisation
- la console pilote
- une unité de disque 80 Mega octets
- 6 consoles dont une console système
- une unité de bandes magnétique 45 IPS, 800/1600 BPI
- une imprimante
- 1 traceur

Le système de visualisation

Le système de visualisation est constitué d'une unité graphique dont les interfaces sont adaptées à l'utilisation de plusieurs types de terminaux de visualisation :

- Tube cathodique monochrome et balayage cavalier
- Tube cathodique couleur à pénétration et balayage cavalier
- Tubes cathodiques couleur shadow Mask et balayage cavalier (tubes utilisés sur avions).

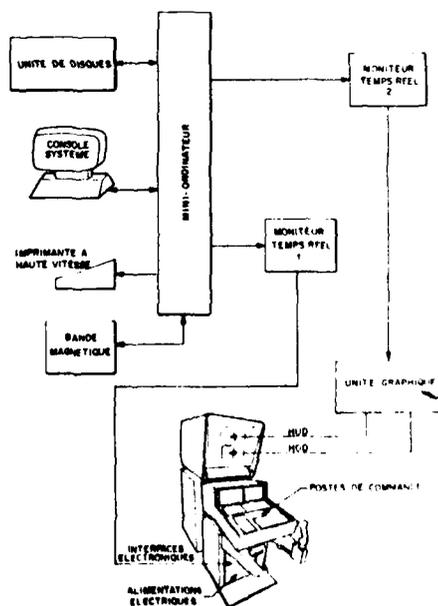
Pour couvrir les besoins de plus en plus nombreux en visualisation de type télévision une étude d'adaptation du système est en cours, pour l'instant des artifices sont utilisés.

L'emploi de l'un ou l'autre des terminaux est déterminé par les besoins propres au type de l'application réalisée, parmi les critères de choix citons :

- le besoin en représentativité du tracé (couleur, qualité du graphisme)
- la dynamique de l'image
- la charge de calcul globale
- qualités propres des différents terminaux

Des artifices matériels et logiciels permettent d'assurer pour chacun de ces couplages une représentativité suffisante pour les besoins de l'étude.

Ensemble schématique



La console pilote

C'est l'organe de dialogue de l'opérateur (pilote, ingénieur) avec la simulation. La machine actuellement en service est représentative des commandes du Mirage 2000 et comprend les organes suivants :

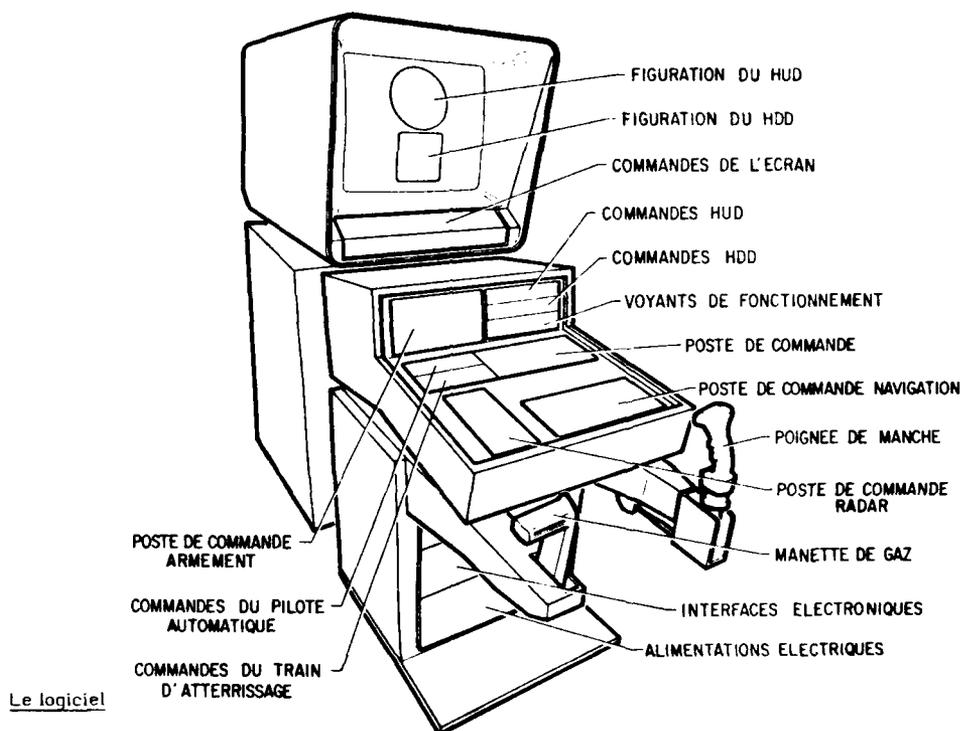
- le manche : celui-ci reproduit exactement, du point de vue de la place des commandes, le manche du Mirage 2000 ; les efforts à exercer sont eux aussi conformes à la réalité
- la manette des gaz : de même que le manche pilote, celle-ci reproduit la manette du Mirage 2000.

Les autres organes de dialogue n'ont pas la prétention de représenter exactement la réalité ; ils sont cependant suffisamment réalistes pour que le pilote s'y retrouve ; la console comprend :

- un PCR (Poste de Commande Radar) présentant les principales sélections :
 - . arrêt, préchauffage, silence, émission
 - . changement d'échelle
 - . PPI / B
 - . angle de balayage
 - . nombre de lignes
 - . PSIC / PSID (poursuite sur information continue ou discontinue)
- un PCA (Poste de Commande Armement) permettant de faire et visualiser les différentes présélections et sélections correspondant aux différentes missions
- un Poste de Commande Navigation (PCN) permettant par exemple d'introduire de nouveaux buts de navigation, de changer de but de navigation, de connaître la position de l'avion, de faire un recalage de la navigation
- un bandeau comportant :
 - . les commandes de sortie et de rentrée du train d'atterrissage
 - . les sélections des divers modes du pilote automatique, à savoir tenue de route, tenue de cap, de pente ...
- une série de clefs ingénieur permettant à un instructeur éventuel :
 - . de simuler des pannes
 - . de modifier certains paramètres : mettre du vent latéral, longitudinal, des turbulences ...
 - . de changer certains types de visualisations : par exemple, on peut préférer voir, au lieu du cercle représentant les limites du champ viseur tête, le dessin de la glace semi réfléchissante sur laquelle est projetée la symbologie
 - . de suspendre la simulation en arrêtant les intégrations, de l'accélérer dans certaines phases peu intéressantes
 - . d'initialiser ou de reinitialiser le système dans diverses configurations : différentes altitudes, différentes vitesses, différentes positions géographiques ; ceci permet de se mettre en situation pour faire rapidement certaines missions par exemple, pour faire une approche, l'avion est initialisé à 1500 pieds d'altitude.

ces clefs peuvent également permettre, lorsqu'il s'agit de mettre au point des systèmes, de choisir différents dessins de réticules, différents gains ...

- des potentiomètres servant par exemple à régler la force du vent, à changer sa direction...



- Structure globale

La fonction remplie par le logiciel au niveau du système OASIS est double : il s'agit d'une part de traiter les signaux provenant de la console pilote et d'autre part de faire tourner le programme de simulation proprement dit. Ces deux tâches sont réalisées par deux programmes :

- programme d'acquisition (ACQ)
- programme de simulation (ESSAI)

Ces deux programmes travaillent en séquence et s'échangent des informations par l'intermédiaire d'une zone mémoire accessible aux deux programmes.

Le programme de simulation

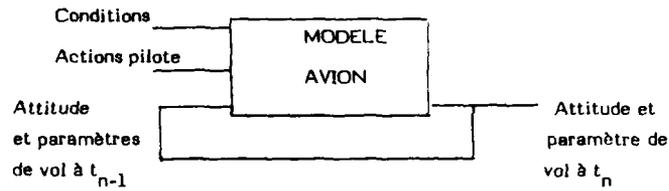
Le but du programme de simulation par lui-même est de simuler le comportement de l'avion et ses réponses aux différentes commandes.

Contrairement au programme d'acquisitions qui est réalisé en ASSEMBLEUR, cette partie est réalisée à 90 % en FORTRAN, pour des raisons évidentes de maintenabilité.

Elle peut se décomposer en deux parties :

- 1 partie temps différé qui effectue certaines initialisations (charge par exemple des tableaux avec les valeurs des sinus et cosinus de degré en degré ; cette tabulation permet de gagner par la suite en temps de cycle), ainsi bien sur le plan calculs que sur le plan des visualisations
- 1 partie temps réel, qui boucle constamment sur elle-même.

La base de la simulation est le modèle avion :



Le modèle utilisé à ce jour est relativement simple, et, malgré des simplifications destinées à améliorer le temps de cycle, il donne satisfaction dans l'ensemble aux pilotes. Ce modèle est bien évidemment appelé à tous les cycles de calcul.

Les nombreuses fonctions différentes du système d'armes ont conduit à scinder la simulation en plusieurs parties (cas du Mirage 2000) :

- les macrofonctions
- la génération des symboles tête haute
- la génération des symboles tête basse
- le logiciel graphique

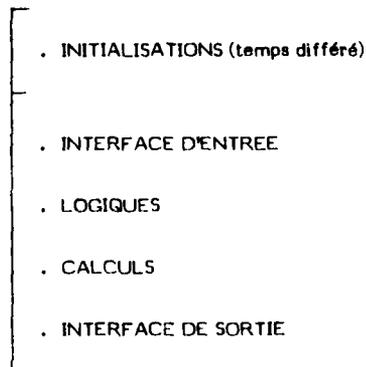
Remarque : pour les applications possédant plusieurs visualisations têtes basses il y a une partie génération de symbole par visualisation.

Les macrofonctions

Elles simulent différentes fonctions du système :

- macrofonction pilotage
- macrofonction navigation
- macrofonction radar
- macrofonctions Air-Sol (une macrofonction par type d'arme)
- macrofonctions Air-Air (une macrofonction par type d'arme)

La structure d'une macrofonction est du type suivant :



Chaque macrofonction est appelée à chaque cycle, et, suivant qu'elle est concernée ou pas (par exemple on peut avoir ou pas un emport de type Air/Sol, si oui la fonction Air-Sol correspondante peut être non

présélectionnée, présélectionnée ...) elle gère sa propre configuration de commande et effectue les calculs, logiques de mode et de présence-absence de réticules correspondant à l'état du système.

Chaque macrofonction est indépendante des autres macrofonctions, certaines sont activées en permanence, d'autres ne le sont qu'en fonction des commandes réalisées par l'opérateur. Les logiques d'activation, images des priorités, des exclusivités, sélection, présélection sont consignées dans des mots de contrôle.

Génération de symbole tête haute :

Cette génération de symbole tête haute joue en gros le rôle du B.G.S. (Boîtier Générateur de Symbole) implanté sur l'avion.

C'est elle qui assure le dessin des réticules du H.U.D. (tête haute).

Ce sous programme est donc une suite de tracés, scindés en sous parties correspondant aux divers réticules.

Il assure une autre fonction du B.G.S. qui est celle du calcul des limitations : certains réticules ne peuvent disparaître du champ viseur, et sont donc "limités", la plupart du temps circulairement, au cercle et représentant le champ du viseur tête haute.

Ce sous programme est parfaitement découplé (ou du moins le plus possible) des différents calculs effectués dans les diverses macrofonctions. Pour la définition de chaque réticule, il n'y a donc qu'un positionnement, les divers paramètres permettant de le tracer (par exemple le roulis commandé dans le cas d'un ordre en roulis), et un bit de présence-absence.

Tracé des réticules

Chaque réticule est représenté en mémoire par un certain nombre de mots. En fonction des bits de présence absence ces mots sont transférés en mémoire de visualisation.

Certains réticules présentent la particularité de ne pas changer de dessin et de rester constamment parallèles à eux-mêmes ; il est inutile, et pénalisant pour le temps de cycle, de retransférer dans ce cas l'intégralité du dessin ; leur description est effectuée une fois pour toutes en initialisation.

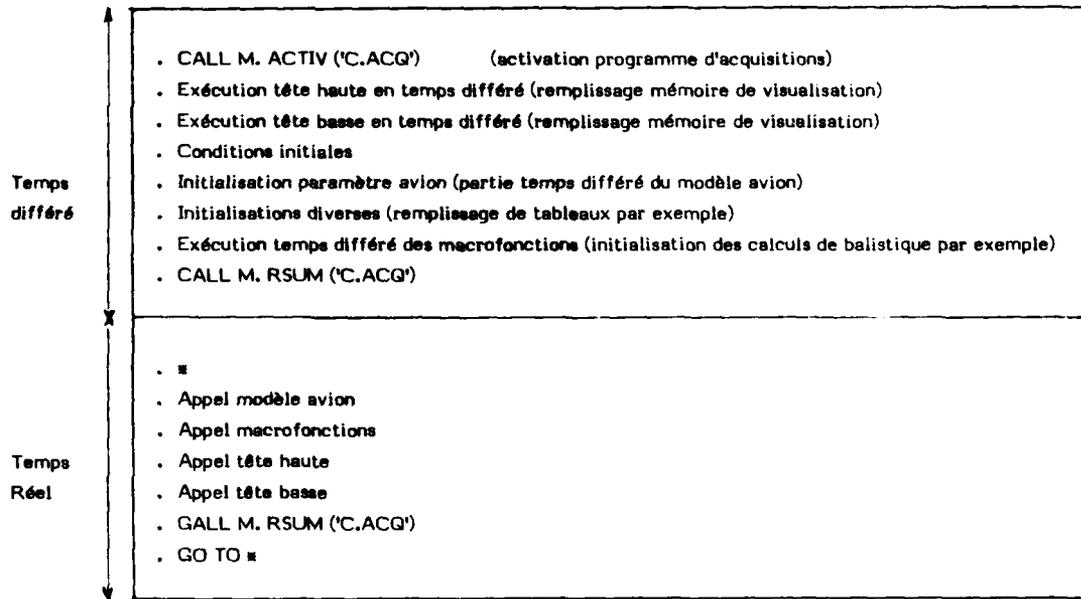
Génération de symbole tête basse

Elle assure le tracé du scope tête basse (HDD) suivant un mode de fonctionnement sensiblement identique à la tête haute (certains réticules sont d'ailleurs identiques en tête haute et en tête basse).

Elle comporte de plus en clôture l'envoi du buffer contenant les codes de visualisation vers l'unité de traitement.

Les limitations sont ici "carrées", donc plus simples.

Synthèse générale de la structure du programme de simulation



Le logiciel graphique de base

Le logiciel assure les tâches suivantes :

- la gestion de la mémoire de visualisation
- les fonctions graphiques élémentaires.

La gestion de la mémoire de visualisation

La mémoire de visualisation est gérée à partir des informations définissant les présences-absences des réticules et les fréquences de rafraîchissement.

Les fonctions graphiques élémentaires

Les matériels du type de ceux utilisés sont pour la plupart livrés avec un logiciel de base permettant de réaliser les fonctions élémentaires de tracé des vecteurs, de positionnement, de tracé de chaînes de caractères, de clignotement ...

Cependant, en ce qui concerne ces matériels, ce logiciel de base ne répond pas à ce qui lui est demandé dans le cadre temps réel, car il a été conçu dans un esprit de Conception Assistée par Ordinateur, où le temps de cycle est un facteur beaucoup moins prépondérant. Il comporte donc beaucoup de tests supplémentaires de cohérence, de contrôle de débordement écran, ... , toutes choses qui font que, lorsque l'on appelle les sous programmes concernés, les chronos s'en ressentent (ces éléments sont en effet appelés plusieurs centaines de fois par cycle).

Ce logiciel de base a été réalisé spécifiquement pour cette application, et consiste en fait à remplir un tableau avec les mots codés que doit recevoir l'unité de traitement, et à gérer l'index du buffer.

Le fait de gérer soi-même le logiciel de base présente de plus un autre intérêt : bien que les mots à envoyer aux divers systèmes graphiques soient différents, il a été possible de réaliser des sous programmes identiques du point de vue entrées/sorties et traitement pour les deux unités graphiques utilisées à ce jour. Il suffit alors, suivant celle que l'on utilise, de spécifier à l'éditeur de liens la librairie qu'il doit utiliser.

UTILISATION DU SYSTEME

Ce système est actuellement utilisé pour réaliser la validation du dialogue "Homme-Machine" des avions d'armes à partir des spécifications globales de chacune des fonctions des systèmes en développement.

Il permet :

- la mise au point des procédures de dialogue
- l'optimisation de la symbologie (forme de réticules, positions)
- l'étude d'algorithmes de guidage.

L'expérience acquise lors de sa mise en oeuvre a conduit à envisager et à réaliser d'autres types de tâches.

- Utilisation pour l'avionique civile

L'apparition des EFIS (Electronic Flight Instruments System) et des FMS (Flight Management System) sur les avions civils a fait ressentir un besoin similaire à celui des avions d'armes.

Le système OASIS a été adapté à ce besoin et est utilisé actuellement pour la définition des symbologies EADI, EHSI, MFD des EFIS des avions d'affaire FALCON, des tubes avions sont utilisés pour cette tâche (voir Photos).

L'étude du dialogue avec un FMS est en cours d'étude avec ce système.

L'utilisation de ce système dans le cadre d'un processus de certification est également envisagé permettant ainsi de limiter le nombre de vols consacrés à la certification de la symbologie.

- Utilisation pour l'instruction

Du fait de sa représentativité au niveau logique système, les utilisateurs de l'avion correspondant peuvent à l'aide de ce système recevoir une formation initiale permettant :

- de limiter le nombre de vols
- de valoriser l'instruction par rapport à la lecture d'un manuel pilote
- d'instruire simultanément plusieurs utilisateurs.

- Utilisation pour la confection de manuels pilote

L'adjonction d'un traceur sur un tel système permet une application à la confection des manuels pilote pour lesquels il est nécessaire de faire des planches représentant les différentes visualisations dans les phases des diverses missions.

Il est possible, grâce au système OASIS de dérouler ces missions puis de figer la situation à un instant donné, ce qui permet de tirer les planches sur le traceur à intervalles réguliers, ou encore aux moments "clef" de la mission, et de reconstituer ainsi les différentes phases sur le papier.

Un autre avantage à cette méthode est d'être sûr que les différents paramètres représentés seront corrélés et cohérents.

- Utilisation pour la spécification détaillée de symbologie

L'utilisation du système dans son rôle de validation de spécifications globales constitue la dernière étape avant la phase de rédaction des spécifications détaillées de symbologie conduisant à la réalisation du logiciel. Du

fait de la représentativité du fichier des réticules celui-ci peut être utilisé pour éditer les spécifications de symbologie à l'aide du traceur.

Les spécifications seront alors conforme à ce qui aura été validé.

CONCLUSION

L'expérience acquise dans l'application du processus d'élaboration du dialogue "Homme-Machine" décrit auparavant s'étend maintenant aux différentes versions de Mirage 2000 et prouve tous les jours le bien fondé de la méthode et des moyens employés ; ceci a conduit à le généraliser à toutes les études de ce type en cours actuellement (ACX, études générales).

En ce qui concerne le système OASIS, sa souplesse d'utilisation, son potentiel d'évolution et la diversité des tâches qu'il permet de réaliser alliés à un investissement initial modéré font de ce système un outil au rapport coût/efficacité très appréciable.

Bien entendu le processus et les moyens décrits ne sauraient résoudre tous les problèmes liés à la mise au point d'un SNA. En particulier tous les problèmes posés par l'avionnage et l'intégration des équipements (réalisation, interfaces) ne sauraient être résolus par le système OASIS et nécessiteront d'autres tâches telles que essais sur banc d'intégration stimulables ou essais sur avion au sol ou en vol et essais sur de gros simulateurs.

TERRAIN FOLLOWING WITHOUT USE OF FORWARD LOOKING SENSORS

By H.F.Schwegler and B.Schreiner

Messerschmitt Bölkow Blohm GmbH
Unternehmensgruppe Hubschrauber
und Flugzeuge
D-8 München 80 Postfach 801160

Summary

A system is presented which allows an aircraft to follow a preplanned vertical profile along a planned route. Using a simple representation of the planned vertical profile instead of terrain data the computational tasks to be performed in the aircraft in real time are minimized. This system may ease some operational problems of today's military aircraft when flying in hostile environment until the advent of more sophisticated systems likely to be based on the intelligent combination of forward looking sensor data and "stored map" derived information.

Introduction:

The advances in defence weaponry, such as radar-laid anti-aircraft-artillery and surface to air missiles, have greatly complicated the task of today's attack aircraft.

In order to maximise the chances of survival in this environment, military aircraft are using terrain-following radars in connection with suitable processors and flight-director or autopilot to provide low flying capability under all weather conditions.

Problem:

The disadvantage of these systems is the considerable effort required to achieve a reasonable jamming resistance with respect to the known countermeasures and in addition the forward radiation may ease the task of hostile intelligence.

These problems may be overcome by using "stored maps" in connection with associated sensors and sophisticated digital processing. Various systems are known, which are operational for missiles and in advanced experimental stages for manned aircraft.

Common to these two systems is the fact that they are associated with fairly high costs and require considerable time for integration and testing.

Solution (or rather a small contribution):

We have implemented in existing operational aircraft a system, which within a restricted envelope, offers the facility to fly low over undulating or hilly terrain without forward or no radiation at all along a preplanned track.

This feature is complementary to the existing terrain following system.

The basic principle of this system is very simple:

A planned vertical profile is stored in the mission computer as a series of altitude values as a function of downrange from an Initial Point.

This series of points is interpolated with a suitable function thus generating a demanded altitude as a function of downrange which is provided by the A/C navigation system. (see fig. 1)

The difference between this demanded and the A/C inertial altitude (updated shortly before or at the Initial Point) basically governs a flight director which is used to fly the A/C at the planned altitude along a planned track.

Details of the implemented system are:

The system altitude used in the vertical flight director calculations is for safety reasons the smaller (i.e. the safer) one of different sources.

If switched on the Radaraltimeter has a monitoring function providing a pull up demand in case of undershooting a preset groundclearance.

The interpolation between the downrange/altitude points is done using a cosine function.

The tasks to be performed in the mission-planning phase are:

1. Generation of the desired flightprofile. This can be done in two ways which are not exclusive:
 - 1.1 Plot terrain profile along planned track using e.g. map data (allowance must be made for an across track navigation error of the system) and draw a desired flight profile allowing for boundary conditions (vertical acceleration, location of potentially known threats) or:
 - 1.2 If a DRMLS (Digital Radar Land Mass) based simulation system is available (e.g. in a Tactical Flight Simulator) a "flight" can be performed using the simulated TF Radar system and/or manual elevation steering under "visual conditions". Recording height and downrange can provide a desired profile.
2. Approximation of the desired profile by a series of points which are connected by cosine functions. (interval 0 - 360 deg) To generate a good approximation for hilly terrain will not require more than 3 to 4 points per 10 nm at most.

3. Verification of the data set by constructing the desired vertical flightpath from the downrange/altitude table using the interpolation function and checking against the available terrain data.

When making up the planned altitude profile the different error sources which dictate the minimum achievable groundclearance (H_{min}) must be considered. The main contributions are:

D_{alt} error of altitude sensor (after updating)
for air data stabilized sensors this will be increasing with downrange.
 D_{bse} error inherent in data base used. These may be maps or a DRLMS database.
 P_{iler} Pilots error for following a flight-director demand
 D_{hor} Horizontal Navigation error along track

These quantities may be related to H_{min} by the following formula:

$$H_{min} > D_{alt} + D_{bse} + P_{iler} + C * D_{hor} * D_{hor}$$

where:

$$C = 0.5 * b_n / (v * v)$$

b_n : maximum vertical acceleration
 v : groundspeed

As the contributions of the different errors are largely independent the sum on the right hand side may be replaced by the "root sum square" of the corresponding N-sigma quantities.

The implementation is using about 300 words of the A/C mission computer and the altitude/downrange data are loaded in the mission-computer in the same way as other mission data.

This small amount of core required is due to the fact that not a terrain but a desired flightprofile is stored. The calculations, which in a terrain following radar system are performed inflight, in our case are practically performed on ground in the flight planning phase.

Limitations:

The system can obviously only be used when a preplanned track is adhered to. This implies a fairly accurate navigation system which in our case was available. Unknown man made obstacles present the same problems as with 'stored map' based systems. The other prerequisites for a simple implementation which were also available are:
Availability of interface with mission computer to 'fill in' the planned altitude profile.
Interface between navigation system, the mission computer and a display to provide the pilot with lateral and vertical steering information.

Results:

Evaluation of flight data shows that the performance of this system is as expected.

Fig. 2 shows 2 'flight profiles' over a terrain profile resulting from simulator trials. The upper one was generated using the terrain following radar and automatic elevation control. The second one was 'flown' using the 'stored profile' approximated by 8 points and using manual elevation control.

Figures 3 and 4 show recorded flight data (planned and actually flown profile)

Conclusion:

The design and implementation of such a system can be done without technical risk and on an economic basis, filling a gap between systems available today and those planned to be available in some years. The system will provide a capability to minimize the jamming, detection and vulnerability of military aircraft when penetrating hostile environment at low level without the use of radiating sensors.

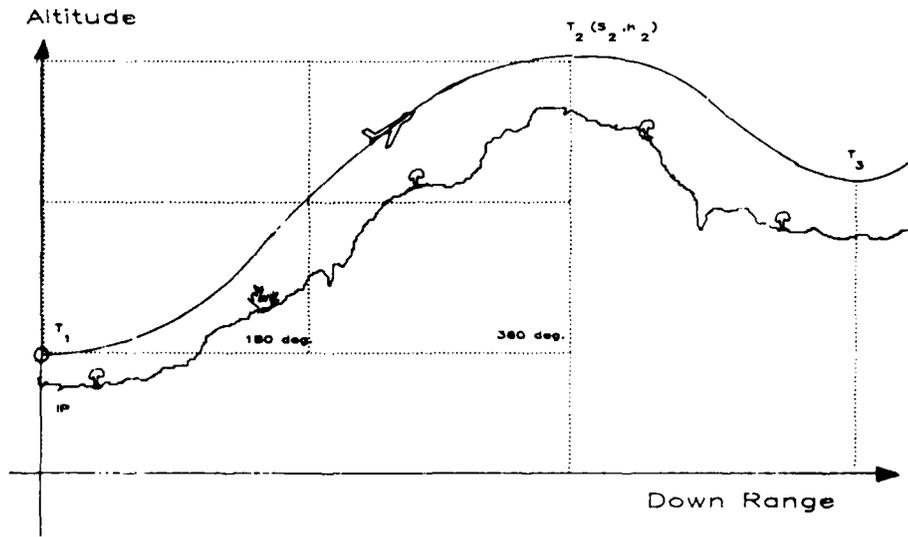


Fig. 1 Profile fitting

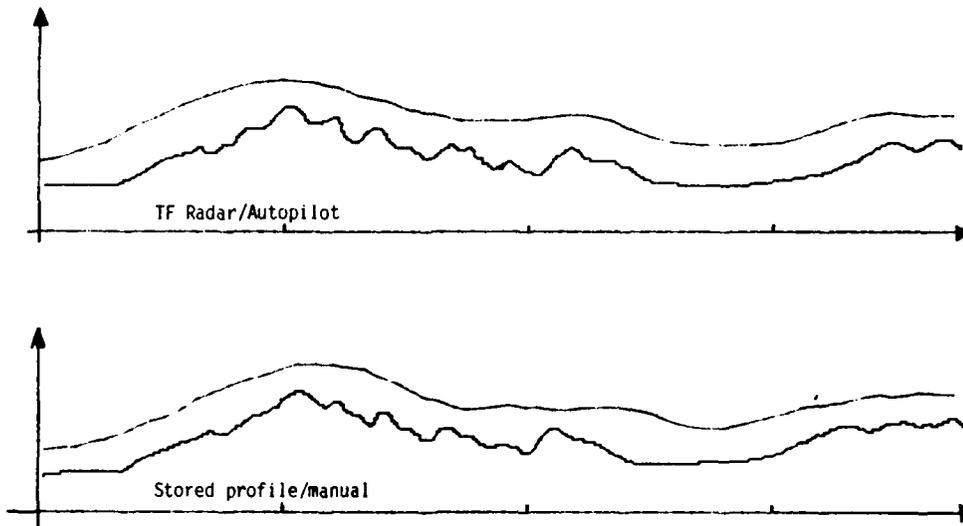


Fig. 2 Simulation data

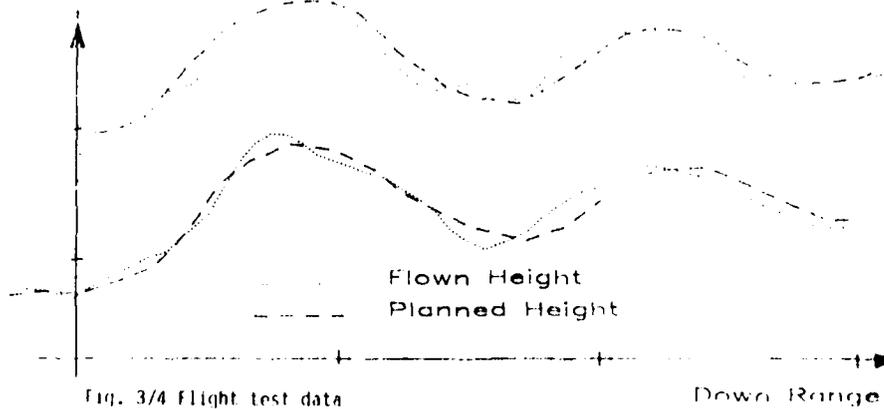


Fig. 3/4 Flight test data

THE USE OF PRESSURE SENSING TAPS ON THE AIRCRAFT WING AS SENSOR FOR FLIGHT CONTROL SYSTEMS

by

Dr.-Ing. Dirk Brunner
 Technische Universität Braunschweig, Hans-Sommer-Straße 66, D-3300 Braunschweig

SUMMARY

For the low speed operation of aircraft, during STOL-take off or STOL-landing and for windshear situations a precise measurement of the state of the aerodynamic flow is required. Normally the dynamic pressure is used to assess the state of flow, thus defining the stall margin in terms of a speed factor. However, flying at higher lift coefficients, a precise maintenance of a given lift coefficient by controlling the speed is no longer feasible. Instead, controlling the angle of attack or controlling the lift coefficient directly should be used.

Some methods for the measurement and the control of the state of the aerodynamical flow including wing tap pressure measurements are discussed. Wind tunnel results are presented, that show the pressure distribution of a slotted STOL-wing and the typical relationship between the tap pressure, angle of attack and flap angle. Wing tap pressure measurements taken with the STOL-aircraft Do 28 are then discussed showing the feasibility of the method described to sense the state of flow.

INTRODUCTION

To maintain a safe stall margin for the low speed operation of aircraft a precise information about the state of the aerodynamic flow is required. Normally, during landing approach this safe stall margin is defined by an approach speed which is 1.3 times the stall speed. However for STOL-aircraft, during wind shear situations and for the maneuvering flight of modern combat aircraft the speed information is no longer sufficient for the determination of the state of the aerodynamic flow. Instead, methods controlling the angle of attack or the lift coefficient directly, should be used.

DETERMINATION OF THE AERODYNAMIC STATE OF FLOW

During stationary flight and near stationary flight in a pullup maneuver the lift L is equal to:

$$(1) \quad L = \frac{W}{n \cdot \cos \phi}$$

where W is the weight, n the loadfactor and ϕ the bank angle. With W^* as the apparent weight

$$(2) \quad W^* = \frac{W}{n \cdot \cos \phi}$$

and the lift as the product of the dynamic pressure q , the wing area S and the lift coefficient c_L :

$$(3) \quad L = q \cdot S \cdot c_L$$

we get as long as L equals W^* :

$$(4) \quad \frac{W^*}{S} = q \cdot c_L \quad \text{or}$$

$$(4 a) \quad c_L = \frac{W^*}{S} \cdot \frac{1}{q}$$

Substituting the dynamic pressure by:

$$(5) \quad q = \frac{\rho}{2} v^2 \quad \text{we get:}$$

$$(6) \quad c_L = \frac{W^*}{S} \cdot \frac{2}{\rho} \cdot \frac{1}{v^2}$$

If we look at the change in the lift coefficient with the change of speed we find /1/:

$$(7) \quad \Delta c_L = \frac{W^*}{S} \cdot \frac{2}{\rho} \cdot \frac{2}{v^3} \cdot \Delta v$$

Assuming that a pilot or an autothrottle system is able to hold the approach speed within ± 5 kts $\approx 2,5$ m/s, at an approach speed of 140 kts ≈ 70 m/s and a wingloading factor W/S of 3 600 N/m² a variation in lift coefficient of about 0,2 due to a speed variation of ± 5 kts results. During a STOL-approach at 65 kts $\approx 32,5$ m/s, the same loadfactor and speed variation as above, a lift coefficient variation of about 1,5 will result. (Refer to Fig. 1). In the low speed regime, where a precise maintenance of the lift

coefficient is needed, the definition of an approach speed will lead to a rather inaccurate control of the aerodynamic state of flow.

For any aircraft there is a transition speed above which speed is the more accurate control parameter and below which the angle of attack or the lift coefficient is the more favourable control parameter. For example, flight tests with a STOL-aircraft, having a wingloading of 3 600 N/m² and an aspect ratio of greater than 4 revealed that the flight control system is capable to control the angle of attack with an accuracy of 0.8° equalling a lift coefficient accuracy of 0.07 /2/. These conditions lead to a transition speed of about 160 kts = 80 m/s. Higher wing loadings will shift the transition speed to higher values. (Fig. 2)

DISCUSSION OF METHODS FOR SENSING THE AERODYNAMIC STATE OF FLOW

Basically to assess and control the aerodynamic state of flow, the parameters:

- speed v
- dynamic pressure q
- the reciprocal value of the dynamic pressure q^{-1}
- angle of attack α and
- lift coefficient

could be used. As shown later, wing tap pressure measurements can be used to determine the lift coefficient directly. To control the aerodynamic state of flow:

- a. the flow state must be measured
- b. the desired flow state must be defined and
- c. the system deviation must be controlled.

For some good reason the parameter speed is used for control as long as a speed has to be maintained for air traffic control or weapon delivery purposes or the structural limits of the aircraft have not to be exceeded. However for:

- best angle climb
- maximum range flight
- safe approach speed and
- steepest decent

it is easier to define a desired flow state in terms of angle of attack or lift coefficient instead of using the flight handbook and evaluate the respective speeds in relation to the actual aircraft weight.

According to Fig. 2 the control of the aerodynamic flow should be accomplished within the low speed regime by control parameters other than speed or dynamic pressure. The typical properties of the different methods to sense and control the flow state are put together in the table below /3/:

| method | | v | q | q^{-1} | c_L | c_L+q | α |
|-------------|------------|----------------|----------------|----------------|----------------|---------|----------------|
| measurement | approach | good | good | good | medium | good | poor |
| | cruise | good | good | good | medium | good | poor |
| | high speed | good | good | medium | unsatisfactory | good | unsatisfactory |
| set-value | approach | poor | poor | poor | good | good | excellent |
| | cruise | poor | medium | medium | good | good | good |
| | high speed | good | excellent | good | unsatisfactory | good | unsatisfactory |
| control | approach | unsatisfactory | unsatisfactory | good | good | good | good |
| | cruise | poor | poor | good | good | good | good |
| | high speed | good | good | unsatisfactory | unsatisfactory | good | unsatisfactory |

Table 1
comparison of different methods for measurement
and control of the flow state

According to the table, using both the dynamic pressure and the lift coefficient for the assessment and control of the aerodynamic state of flow will grant the best overall performance of the flight control system.

DIRECT MEASUREMENT OF THE LIFT COEFFICIENT

In a wind tunnel it is a quite common procedure to measure the pressure distribution along an airfoil and to determine then by integration the lift coefficient. With a STOL-aircraft Do 28 this method was applied to determine the lift coefficient during flight. To get sensor costs down it is desirable to sense the wing pressure at just one discrete position. As long as there is a symmetrical airfoil with a fixed centre of pressure

- the pressure distribution will be proportional to the dynamic pressure and
- the pressure distribution will be geometrically similar when the angle of attack is changed.

Knowing at a reference state of flow the amount of lift L_0 , the wing area S , and the tap pressure p_{m0} , it is possible to determine the lift coefficient by sensing the tap pressure p_m and the dynamic pressure q using the relationship:

$$(8) \quad c_L = \frac{L_0}{S p_{m0}} \cdot \frac{p_m}{q}$$

So for an airfoil with a fixed pressure centre the lift coefficient is directly proportional to the quotient of tap pressure p_m and the dynamic pressure q .

The aircraft used for the tap pressure measurements was a Do 28 D 1 STOL aircraft (Fig. 3). The wing is equipped with a fixed slat and double slotted landingflaps. To optimize the position of the pressure tap it was very useful to have wind tunnel results /4/ at hand which show the pressure distribution of an airfoil that is very similar to that of the Do 28-aircraft. Fig. 4 shows the pressure distributions at 0° landingflaps and Fig. 5 at 30° landingflaps. At the upper side of the wing the pressure maximum is situated constantly at 17% wing cord and - even as this airfoil is not a symmetrical one - the pressure distribution appears geometrically similar when the angle of attack is changed. Referring to Fig. 4 and Fig. 5 it is quite clear to install the pressure sensing taps only at the upper side of the wing as at the lower side the pressure maxima do not show a unique tendency and the pressure gradients to be expected appear very low.

In order to find out the optimum position for the pressure tap at the Do 28 aircraft three different positions for the pressure tap installation were chosen. Fig. 6 shows these positions. In the wing span direction the positions were defined so as to minimize the influence of the aileron and of the propeller area. In the longitudinal direction the wing spar limited the most forward tap position to 25% of the wing cord.

The tap pressure was routed from the tap to a differential pressure transducer, sensing the static pressure minus the tap pressure.

FLIGHT TEST RESULTS

The flight tests were conducted with a Do 28 D 1 STOL-aircraft (Fig. 3). The aircraft is equipped with a nose-boom-mounted angle of attack and angle of sideslip sensor. The static and pitot pressure sources are mounted on the top of the rudder fin. The aircraft has an extensive sensor equipment allowing to record 32 flight mechanical parameters on a magnetic tape unit that stores these data in a pulse coded format at 92 Hertz. Some of the test results shown were plotted online during the test flight on an X-Y-plotter.

The first flight test objective was to determine to which extent the tap pressure is proportional to the dynamic pressure. Beginning at 135 kts \approx 67 m/s and with a constant power setting at 40% rated power the aircraft was slowly decelerated to about 50 kts \approx 25 m/s. The deceleration was achieved by a slow change of the flight path angle so that the stationary flight requirement $W^* = W$ was satisfied. Fig. 7 and 8 show the tap pressure p_m versus dynamic pressure q for the tap No. 1 and 2 respectively. At tap No. 2 there is only a very small variation of tap pressure with the flap setting. As the tendency changes at tap No. 1 (- there, a deflection of the landing flaps reduces the tap pressure-) a tap that would be positioned slightly forward of tap No. 2 should have no change of tap pressure due to flap movement. The results of tap No. 3 are not discussed here as the pressure gradient is smaller than at tap No. 1 and 2 and there are no advantages at that tap position. All in all the linearity of the $p_m - q$ relationship was found to be good.

The next objective was to determine the $p_m - \alpha$ angle-of-attack relationship. Fig. 9 and 10 depict the respective plots. The flight test conditions were the same as above. Again the linearity is very good. At 52° flaps and 11° of angle of attack there is a discontinuity of the tap pressure which corresponds to a change of angle of attack of about 1.5°. However as an angle of attack of that magnitude is only reached in the landing flare, this effect should not limit the use of the tap pressure signal in flight control systems.

As these flight tests prove the linear relationship between the angle of attack, the flap angle and the quotient of tap pressure and dynamic pressure, the tap pressure measurement can provide a signal for flight control systems representing either the angle of attack or the lift coefficient.

The expected good dynamic qualities of the pressure tap signal were confirmed in the flight test. Fig. 11 is a plot of a rapid vertical maneuver resulting in a 0,7 g-load. The maximum difference between the angle of attack sensor α_{Ref} and the tap-pressure-derived signal α_p is below 2 degrees (refer to the $\alpha_{Ref} - \alpha_p$ plot, fig. 11). The sensitivity of the tap pressure transducer to vertical acceleration might contribute to the error. For gust alleviation systems a pressure tap sensor could be advantageous as it senses the higher frequent portion of the horizontal - and vertical gusts /3/.

One disadvantage of the tap pressure measurement is the sensitivity to the errors of the static pressure system. Normally the pressure tap signal, that is fed to the flight control system, consists of:

$$(9) \quad \frac{P_m}{q} = \frac{P_s - P_{tap}}{P_{pitot} - P_s}$$

For the conditions of the D0 28 test aircraft at:

$$\begin{aligned} 2000 \text{ ft} &\approx 600 \text{ m altitude} \\ 100 \text{ kts} &\approx 50 \text{ m/s airspeed and} \\ 7^\circ &\text{ angle of attack,} \end{aligned}$$

a pressure coefficient of 1.28 will result. A sideslip angle of about 7° will cause a static pressure error of -0.5 mb , changing the tap pressure signal to 1.21 which in turn causes a 1° -error in the angle of attack. Fig. 12 shows the flight test results at sideslip angles for the 0° flap setting.

In Fig. 9 the change of the tap pressure due to engine power changes is shown. The lower curve for $\eta_k = 0$ represents about 40 % of rated power where as the upper curve corresponds to about 60 % of rated power causing a 0.5° error of angle of attack. η_k is the angle of landing flap deflection.

To give an overall impression about the signal quality of tap pressure measurements, during flight tests, the tap pressure derived signals are compared with the reference signals in Fig. 13 and 14. In Fig. 13 the reference signal for the angle of attack was the noseboom-mounted angle of attack sensor. When taking the data for the higher angle of attack portion, gustiness might have contributed to the higher dispersion.

For stationary horizontal flights it is possible to compute the lift coefficient c_l according to equation (4a) with $W^* = W$. These computed c_l -values are compared with the tap pressure derived c_l -values as shown in Fig. 14. The dispersion in the c_l -values ranges in the order of 0.2 providing a useful signal for flight control systems.

CONCLUSION

The flight test data show the usefulness of a pressure tap signal for flight control systems. Selecting the proper tap pressure position will greatly influence the success of tap pressure measurements. As the pressure gradient at 25 %-cord for the tap No. 1 and 38 %-cord for the tap No. 2 is relatively low, a high parameter sensitivity results in respect to the tap pressure parameters.

Further examinations should focus on the following points:

- optimum position of the tap
- methods for error compensation
- new pressure sensors (pressure pill)
- redundant sensors
- flight in icing conditions
- transferability of results.

REFERENCES

- 1 Schänzer, G. Verbesserung von Flugleistungen und Flugeigenschaften von STOL-Flugzeugen durch das Flugführungssystem FRG 70 - GCU 70
Bodenseewerk Gerätetechnik GmbH, TB 000 D 759/72, Dez. 1972
- 2 Böhret, H. Realisierung des integrierten Flugregelungssystems FRG 70 im Flugversuch
Bodenseewerk Gerätetechnik GmbH, TB 000 D 774/72, Dez. 1972
- 3 Schänzer, G. Ermittlung des aerodynamischen Strömungszustandes und des Fluggewichts aus Flügel-druckmessungen
Bodenseewerk Gerätetechnik GmbH, TB 000 D 818/73, Okt. 1973
- 4 - Windkanaluntersuchungen an einem Flügelprofil mit Vorflügel und Doppelspaltklappe
Dornier WK Bericht Nr. B 2/57, 1957.

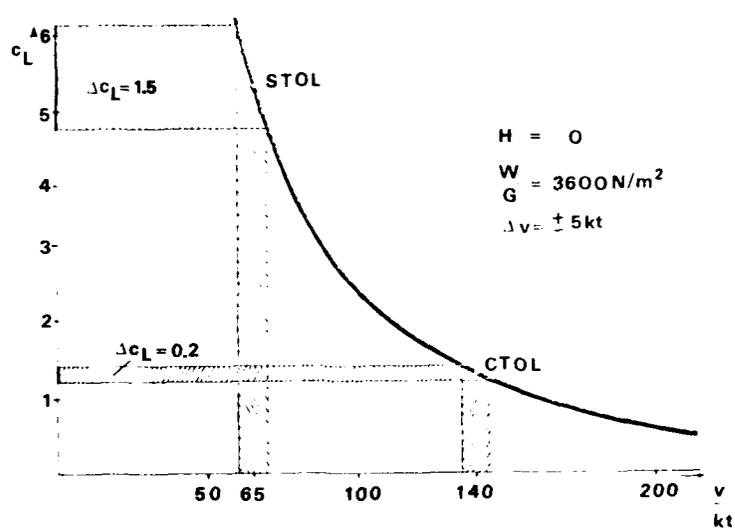


Fig. 1 The influence of the speed error on the lift coefficient c_L .

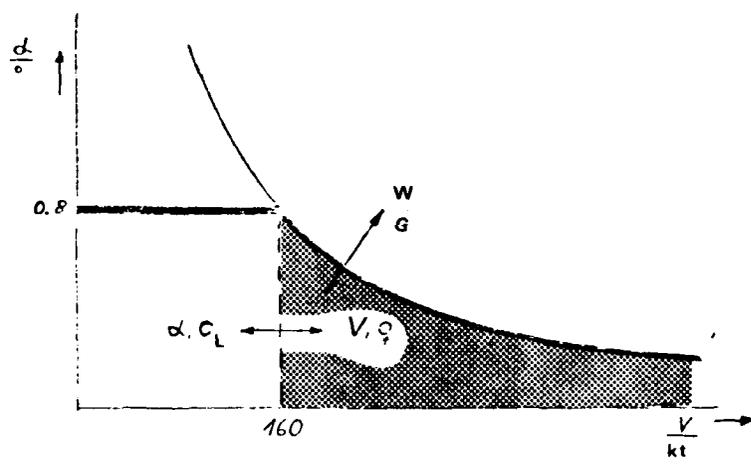


Fig. 2 Favourable regions for angle of attack and speed-control.

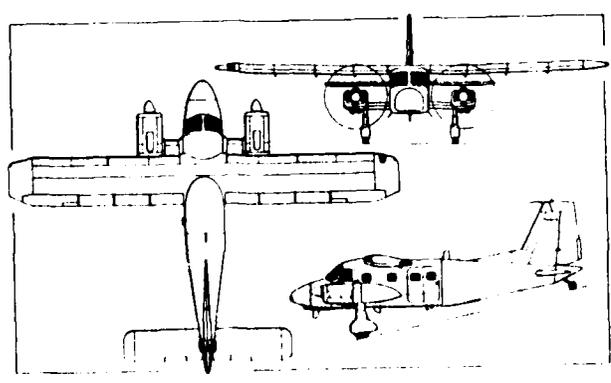


Fig. 3 The DO 28 aircraft

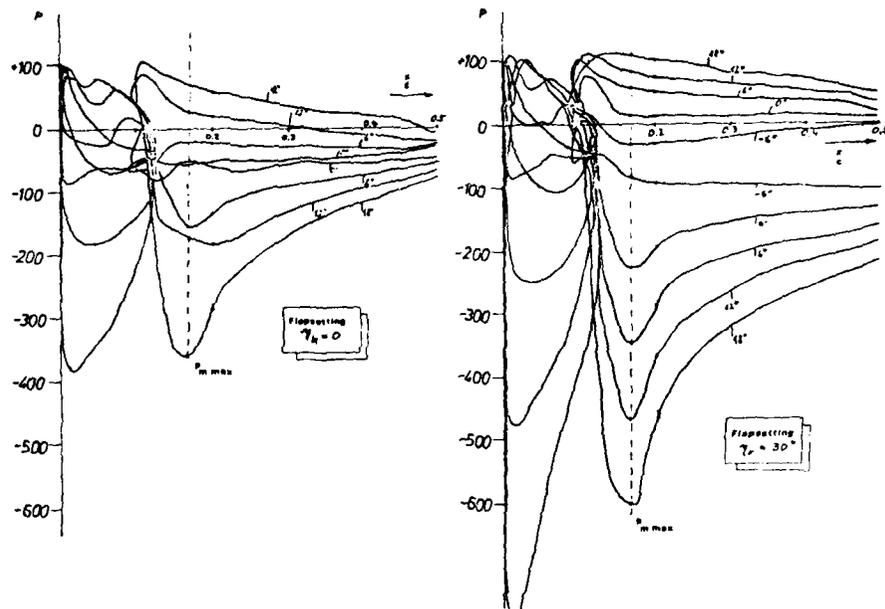


Fig. 4 and 5: Pressure distributions at a wing with slat and slotted flaps:

0° flaps.

30° flaps.

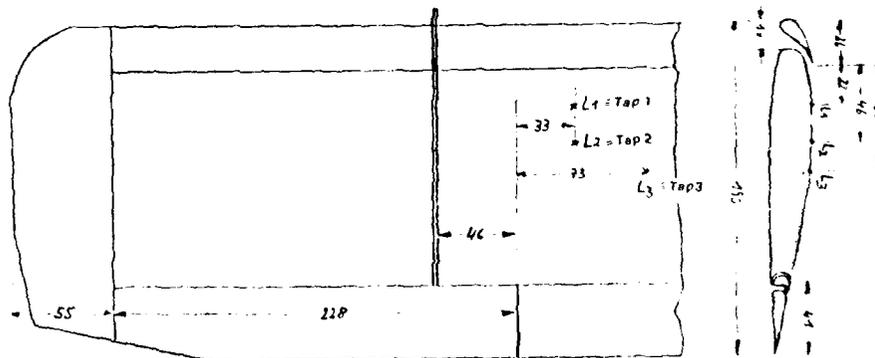


Fig. 6 Position of the pressure taps.

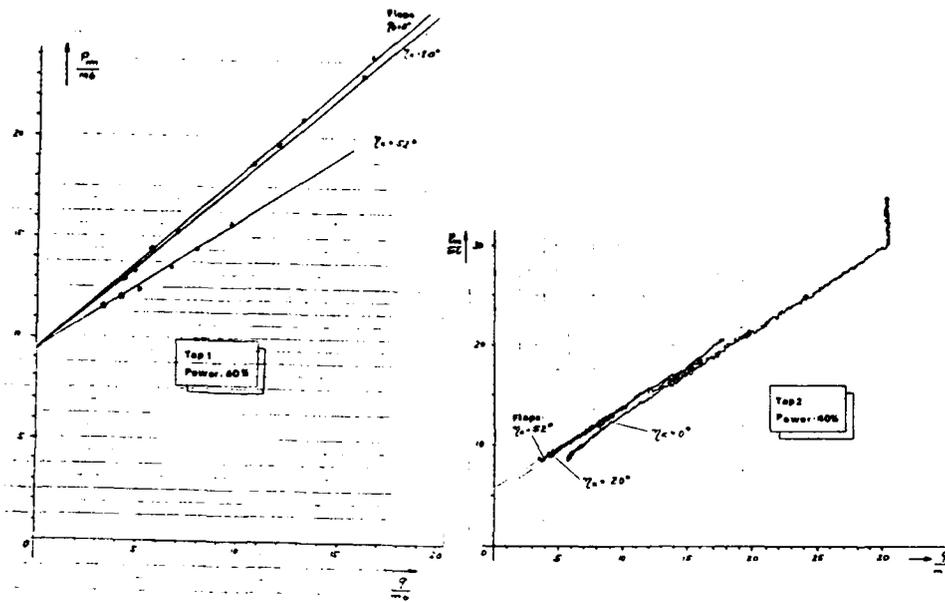


Fig. 7 and 8: Wing pressure versus dynamic pressure:
tap 1, tap 2.

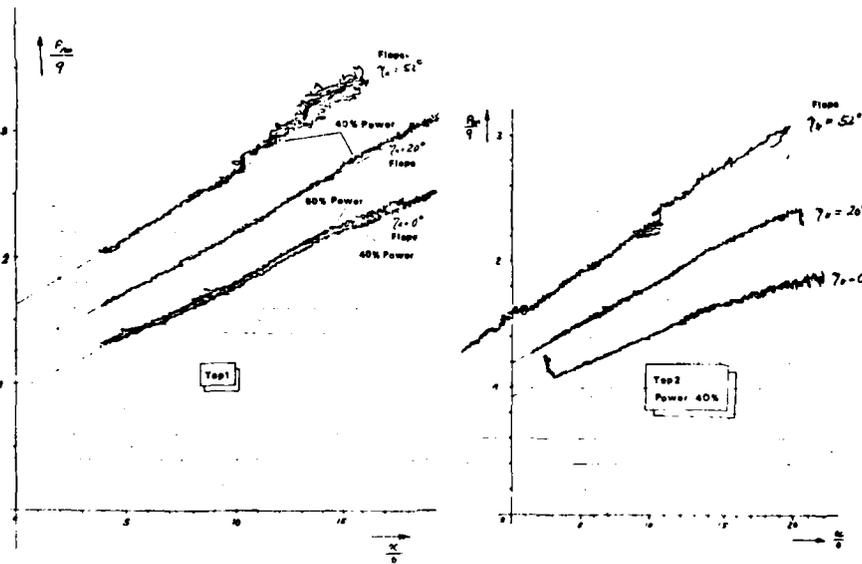


Fig. 9 and 10: Quotient of wing pressure to dynamic pressure p_m/q versus angle of attack:
tap 1, tap 2.

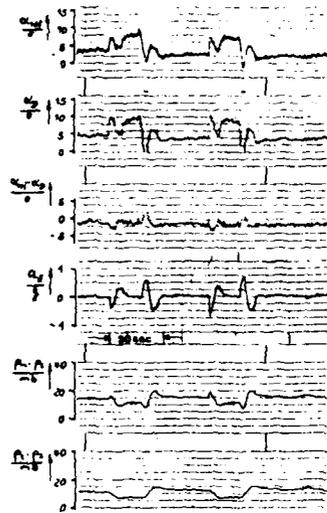


Fig. 11 Dynamic response to a 0.7 g-vertical acceleration.

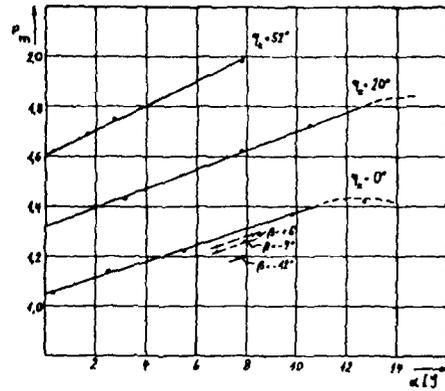


Fig. 12 Effect of sideslip on the wing pressure

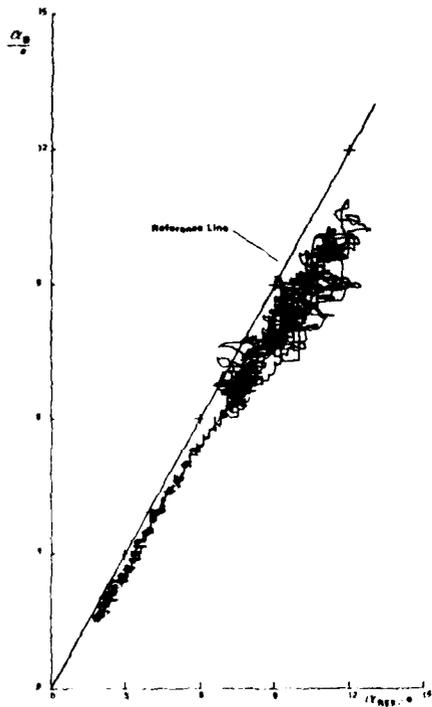


Fig. 13 Dispersion of the wing pressure derived angle of attack signal α_D .

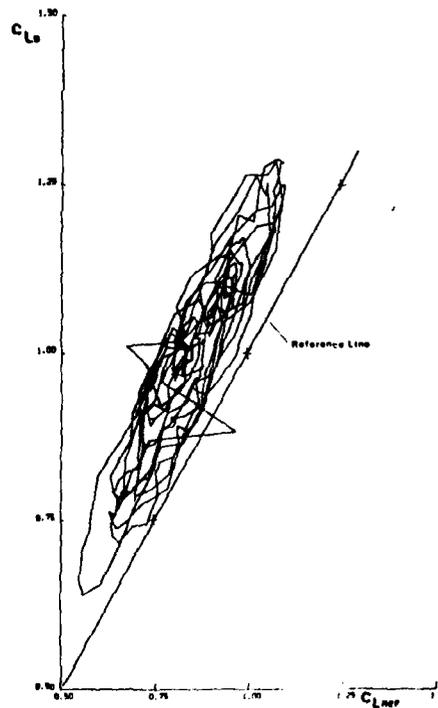


Fig. 14 Dispersion of the wing pressure derived lift coefficient signal c_L .

LOW COST TWO GIMBAL INERTIAL PLATFORM AND ITS SYSTEM INTEGRATION

R.N. PRIESTLEY and J.B. TOWLER
 FERRANTI DEFENCE SYSTEMS LTD
 NAVIGATION SYSTEMS DEPARTMENT
 SILVERKNOWES, FERRY ROAD, EDINBURGH EH4 4AD

SUMMARY

The development of more accurate inertial systems has led to more expensive inertial instruments and complex electronics. Future navigation systems will require further improvements in accelerometer and gyro design in order to achieve the required performance. There are however doubts as to whether autonomous inertial systems will be sufficiently accurate.

An alternative approach is to combine a second sensor with the inertial system to improve performance, for example doppler velocities or NAVSTAR/GPS positions and velocities. The inertial platform may then be simplified, retaining high accuracy only in the platform parameters having an impact on the integrated system performance. This simplification allows lower cost of ownership with an increase in reliability and a medium accuracy reversion capability. The inertial system computer may be programmed with the integration software so avoiding a separate dedicated integration computer with the consequent decrease in reliability.

Introduction

During the last 20 years inertial technology has been developed and applied to resolve the problems of airborne navigation, both civil and military. During this period there has been a shift from single axis floated gyros, through two axes dry gyros to laser gyros used in strapdown configurations with parallel and significant improvements in accelerometers. These developments have been driven by the need to improve accuracy, reliability and lower cost of ownership. At this stage the civil operators appear satisfied that a triple inertial system meets their requirements in terms of accuracy and reliability and apparently is cost effective. However, the military user of high performance fixed wing and rotary wing aircraft have more exacting requirements for position and velocity accuracies to be achieved in a much more demanding flight profile. Doubt exists whether the desired improvements can be met cost effectively by the costly design and manufacture of still more accurate gyros and accelerometers. It may be that a more modest inertial system adopted for velocity or position updating by other sensors may offer significant benefits in accuracy and cost.

Two Gimbal Inertial System Salient Features

The Ferranti FIN 1110 (2-GINS) is a single LRU 2-gimbal schuler tuned inertial navigation system. It has been designed for best performance at lowest cost for operation in any vehicle not required to perform complete rolls or loops directly through the vertical. It is capable of manoeuvre up to and ten degrees past the vertical and is suitable for operation in the most agile helicopters and in most non-combat and transport aircraft.

Costs have been reduced largely by accepting the one constraint of limited freedom in Pitch and Roll. This, together with the adoption of strapdown operation in Azimuth, has allowed the use of only two gimbals instead of four. Further savings have been effected by employing unique design techniques to combine the functions of gimbal drive motor and angular sensor into one unit.

The isolation of the instrument platform from pitch and roll rates allows the use of a high quality two axis "Oscillogyro" DTG for the two vertical channels.

A very high standard of performance is maintained by fitting three high grade FA2 accelerometers to match the Oscillogyro accuracy. For strapdown operation in the azimuth channel a GG1111 floated rate gyro provides good performance at low cost.

The use of the Oscillogyro in the two vertical channels enables the 2-GINS to determine initial True North within ± 0.1 degree.

Since the instrument platform is maintained stable in the horizontal plane the full performance of the FA2 accelerometers is available for the navigation calculations with no degradation due to coupled components of the Earth's gravitational field.

Reliability is increased not only by the reduction of 2 gimbals but also by the resultant reduction in the number of servo amplifiers and their associated electronics. The smaller platform assembly is more rugged and the use of tapes instead of sliprings gives a further gain in dependability.

Although possessing a useful pure IN capability the 2-GINS is primarily intended to be used in "mixed" systems in which the inherent advantages of an inertially based device in terms of accurate gyrocompass alignment, heading and attitude are combined with accurate velocity or position derived from some other devices or positioning systems. The mixing between the two (or more) contributing systems is accomplished within the 2-GINS using the Kalman filter incorporated in the 2-GINS computer software.

Suitable contributory inputs to the Kalman filter can be derived from the following sources:

| | | | |
|-----|------------------|---|----------|
| (a) | Doppler | - | Velocity |
| (b) | Navstar | - | Position |
| (c) | Tacan | - | Position |
| (d) | Omega | - | Position |
| (e) | Tercom | - | Position |
| (f) | Air Data system | - | Velocity |
| (g) | Vehicle Odometer | - | Velocity |

For helicopter operations Doppler velocity is by far the most important source at the present time, although GPS position may become significant in the future.

A Doppler/2-GINS mix offers the following advantages:

- enhanced heading accuracy matches Doppler velocity accuracies
- elimination of magnetic compass vagaries
- rapid alignment techniques allow take off reaction time as low as 30 secs (alignment completed in the air)
- in-air alignment can give full performance accuracies
- ship borne operations with full performance accuracies

In addition to these advantages over Doppler/Flux Valve or pure IN system the 2-GINS has a sufficiently good pure IN performance of its own to allow the Doppler to be used in a "burst-aided" mode wherein it is switched on at intervals in short bursts to provide corrective inputs to the Kalman filter. This mode has obvious advantages in sensitive situations requiring minimum radio or radar transmissions.

Integration with Doppler Radar

The FIN 1110 is a medium accuracy inertial navigation system which is capable of accurate gyrocompassing. Doppler radar provides accurate velocities but with significant short term noise. However accurate azimuth is required for navigation which may only be provided by an inertial system. The good long term performance of doppler radar is common to most radio navigation systems which have noisy short term outputs. This noise approximates to a gaussian random variable and so the doppler and inertial data may be combined in a Kalman filter.

The FIN 1110 Kalman filter estimates the long term errors in the inertial instruments such as platform tilts or misalignments and gyro drifts. The filter also estimates scale factor and beam misalignment errors in the doppler radar. The short term noise in the doppler outputs is also effectively filtered.

The FIN 1110 filter estimates 13 inertial system error parameters and two doppler radar errors. The filter software is combined with the normal navigation calculations and is iterated at 7 - 8 seconds. Since the filter is estimating slowly changing errors this iteration rate is easily adequate. The total program size of the filter is 13 Kbytes and 2.5 Kbytes of random access memory is used. The filter operates open-loop only correcting the FIN 1110 outputs. However after long periods and during initial flight after a poor alignment the errors in the FIN 1110 may increase to large values. In this case the filter outputs will be used to periodically rapidly re-level the inertial platform. This technique has been proven on other Ferranti systems and eliminates many potential problems associated with continuous correction of the platform errors.

The continuous estimation process in the Kalman filter allows a poor initial alignment of the FIN 1110 with the alignment being completed after take-off.

Modes of Operation in Helicopters

There are two distinct environments to be considered - landbased and shipborne. Of these two it is the shipborne application that poses the more significant problems.

Shipborne Operation Alignment

The alignment of an aircraft pure INS on board a ship is usually executed by employing techniques which "transfer" the alignment of the ships SINS or a dedicated transfer INS platform to the aircraft INS. This method is not cost effective for helicopter operations and is unsuitable for most ships other than large carriers.

The advantage of a mixed IN system such as Doppler/IN is that the alignment on board the ship need not be a very high quality. As soon as the helicopter is airborne and the doppler inputs become valid the Kalman filter begins the task of estimating the system errors and refining the alignment. Even so it is still necessary to perform an initial on-deck alignment.

The problem of obtaining the best quality initial alignment before take-off is simply one of finding or inserting a best estimate of true heading and of determining platform horizontal as accurately as possible whilst subjected to the ships motion in terms of pitch, roll, yaw, heave, sway and surge. In this environment the unique 2-gimbal construction of the FIN 1110 2-GINS has an advantage over a full strapdown system as the sensor instruments are maintained physically horizontal and are isolated from the effects of pitch and roll i.e. from the two most severe parameters of ship's motion. It is therefore possible for the Oscillogyro to attempt a direct gyrocompassing operation to define the axis of the earth's deg/hr rotation rate without this being obscured by the deg/sec rates of ship's roll and pitch. However the penalty attached to this operation is that the alignment time constant has to be increased to a value at which the alignment process is relatively unaffected by ship's motion. In operational usage this usually results in an unacceptably long on-deck alignment time. Fast shipborne alignment can be achieved by slaving the 2-GINS heading to the magnetic compass and performing a simple on-deck levelling procedure. Although the magnetic compass information can be very inaccurate on-deck it improves immediately the helicopter has lifted off. At the same time the doppler velocities become valid and the fine levelling can be completed within a few minutes of

take-off. As soon as the pilot is satisfied with his attitude and magnetic heading information he can switch into the Doppler/IN/Filter self gyrocompassing navigation mode. Figure 1 indicates typical Airborne alignment times.

In the intervening period between take off and the time when specified performance is achieved there will inevitably be a build up of position error. This error will be inversely proportional to the quality of the initial alignment and to the speed of operation of the Kalman filter.

Operational Modes

"NORMAL" Navigation Mode

This is a doppler/IN mode performing dead reckoning navigation using the refined doppler velocities and IN heading available from the Kalman filter. Accurate attitude and true heading are available for output to other aircraft systems.

The doppler radar may be operated continuously or in a "burst-aided" mode (BAM) with a duty cycle as low as 1 to 10.

Reversionary "pure-inertial" Mode

At times when the doppler signal is absent or becomes "invalid" the navigation process is continued with the 2-GINS operating as a pure inertial navigator.

The performance accuracies in this mode are very much dependent on the degree of refinement of the error estimates in the Kalman filter at the time of switch over. If the Kalman filter has converged to its steady state limits then a position performance of in the order of 1 nm for the first hour can be expected.

"FAST-ERECT AHRS" Mode

Fast erection can be performed on the ground or on board ship and also in the air during static hover or straight unaccelerated flight.

One of the advantageous aspects of Oscillogyro performance is that it enables the design of a very fast capture and alignment loop typically less than 30 seconds.

In this fast erection mode the 2-GINS has its azimuth output slaved to magnetic compass heading and is only required to level its platform to the horizontal to provide for attitude outputs. After erection the system can be switched into an AHRS mode or into the NORMAL doppler/IN filter mode or into a Schuler tuned "AHRS" mode which allows the aircraft freedom for normal manoeuvre. In the AHRS mode heading remains slaved to magnetic compass (plus variation).

Landbased Operation

This is really a simplified case of shipborne operation with the one difference that the normal alignment method will be 5 min gyro-compassing with automatic transition to doppler/IN filter operation after wheels off and at the appearance of a doppler valid signal. Typical alignment time is indicated at figure 2.

Land-Sea Switching

Doppler radar produces different results over water from those experienced over land. Over water the doppler suffers a scale factor error which is predictable and a bias error caused by surface water movement which is unpredictable. In addition to these errors it is found that the doppler signal noise characteristics are also changed.

In most doppler radar navigation systems a manual Land/Sea switch is provided to enable scale factor error compensation to be effected at the Land/Sea interfaces.

With a Doppler/Flux Valve navigation system any delay in implementing the Land/Sea switching will result in a position error proportional to the change in scale factor x velocity x time delay in effecting the Land/Sea switch. There are no other effects.

In an integrated Doppler/Kalman filter/INS system an unswitched Land/Sea crossing will introduce a sudden velocity error into the Kalman filter which it will interpret as a step change in estimated azimuth error. Unless this situation is quickly corrected by the proper use of Land/Sea switch the Kalman filter state estimates will suffer a significant disturbance which will take several minutes to decay. Computer simulations of this phenomenon indicate that the disturbance can result in considerable position errors.

The Ferranti 2-GINS continuously monitors the difference between the raw doppler velocity data and the inertially smoothed outputs of velocity from the Kalman filter. Any sudden velocity error such as that caused by a Land/Sea crossing is detected. A flagged warning is then presented to the pilot, the Kalman filter state elements are frozen and the navigation system automatically put into the free inertial mode. To return to normal Doppler/IN navigation the pilot must negate the flag either by implementing the appropriate Land/Sea switching or else overriding the warning flag.

Figure 3 gives an indication of the expected relationship between heading error and the error in Doppler velocity.

FIN 1110/Doppler System Accuracies

The initial alignment of the FIN 1110 is 0.15° r.m.s. within 2 - 5 minutes of swith-on. Operating in a continually gyro-compassing mode it accepts the filtered velocity of the doppler, and the accuracy of the heading during flight is a function of the accuracy of the doppler velocity as indicated in Figure 3.

The convergence of the filter and, hence, navigation accuracy, depends on time, flight profile, quality of the ground alignment and quality of the aiding input. It is impossible to accurately specify the performance for every situation

in this paper. Simulation software describes the performance of the FIN 1110 equipment given a flight profile and aiding input. This software may be run to evaluate the FIN 1110 against specific requirements and plots made available, showing azimuth position and velocity accuracies for given situations. Figures 4.1 - 4.6 illustrate a typical flight profile of a tactical helicopter and indicate expected accuracies of north and east velocities, azimuth, doppler along track scale vector error and doppler misalignment angle.

The convergence time of the Kalman Filter is a function of the availability and quality of the Doppler or GPS information but is typically 30 minutes for full convergence.

Loss of Doppler or GPS information shall cause the FIN 1110 system to revert to a pure inertial NAV mode. The secondary navigation performance shall range from 1 nm/hr r.m.s. to the autonomous performance level of 6 nm in 1 hour depending on the convergence of the FIN 1110 Kalman Filter.

Illustrations 5 A/B illustrate flight accuracies achieved on a complex route (typical of tactical helicopter) using a Decca 17 hyperbolic chain with a one sigma accuracy of 200 metres as the datum reference. This datum is not really sufficiently accurate for the purpose, but indicates an overall accuracy of the IN/doppler system significantly better than 1 nautical mile. It is noteworthy that this accuracy is maintained over the whole 90 minute flight.

FIN 1110 with NAVSTAR/GPS

In situations where doppler combined with the FIN 1110 is not adequate NAVSTAR/GPS is likely to be a more accurate alternative.

The NAVSTAR/GPS satellite navigation system provides high accuracy position and velocity to any receiver with access to the high accuracy 'P' code. The NAVSTAR receiver outputs may be combined with the inertial data in the FIN 1110 Kalman filter. The filter may accept position and velocities in geographic axes or the range and range rate to each satellite. This information is processed to form low noise high accuracy position and velocity and estimates of the FIN 1110 azimuth error and gyro drifts. Figure 6 indicates a possible System Configuration.

The FIN 1110 may provide velocities in geographic axes in order to aid the NAVSTAR receiver tracking loops. The bandwidths of the tracking loops may then be reduced increasing the anti-jam performance. The Kalman filter will predict the errors in the FIN 1110 allowing high accuracy navigation when the receiver is jammed or loses signals due to aerial shadowing. This error prediction also results in faster re-acquisition of the NAVSTAR signal.

The NAVSTAR receiver expects the velocity aiding in geographic axes. Extensive simulation has shown that small inertial system azimuth errors may result in large aiding errors during manoeuvres. The continuous gyro-compassing of the FIN 1110 aided by NAVSTAR position and velocity should minimise this effect.

The use of NAVSTAR allows rapid initial gyrocompassing. Initial platform levelling may be completed in 30 seconds and with a magnetic input to provide an initial azimuth the FIN 1110 may become airborne. The correct azimuth would then be determined during flight. The Kalman filter will automatically remove the accumulated position error caused by the poor initial azimuth value.

Conclusions

Pure inertial systems, no matter how accurate suffer from decrease in accuracy with time and must therefore be bounded by some other sensor where very high position and velocity accuracies are required.

Doppler by virtue of good long term velocity combines well with relatively low grade inertial. At the present time, and at today's costs, it would appear that a doppler-inertial mixed system such as has been described offers a significantly cheaper and more accurate solution to navigation and weapon aiming than pure inertial with good reversionary capability. However, Doppler is a radiating sensor and continuous emission is not desirable in the military environment. Reduction of Tx power and burst aiding of inertial may be acceptable.

When GPS becomes fully operational in 1988/89 integration with a low cost inertial sensor promises high position accuracy, accurate attitude and velocity information for enhanced weapon delivery under all conditions of battlefield jamming environment, all at an affordable cost.

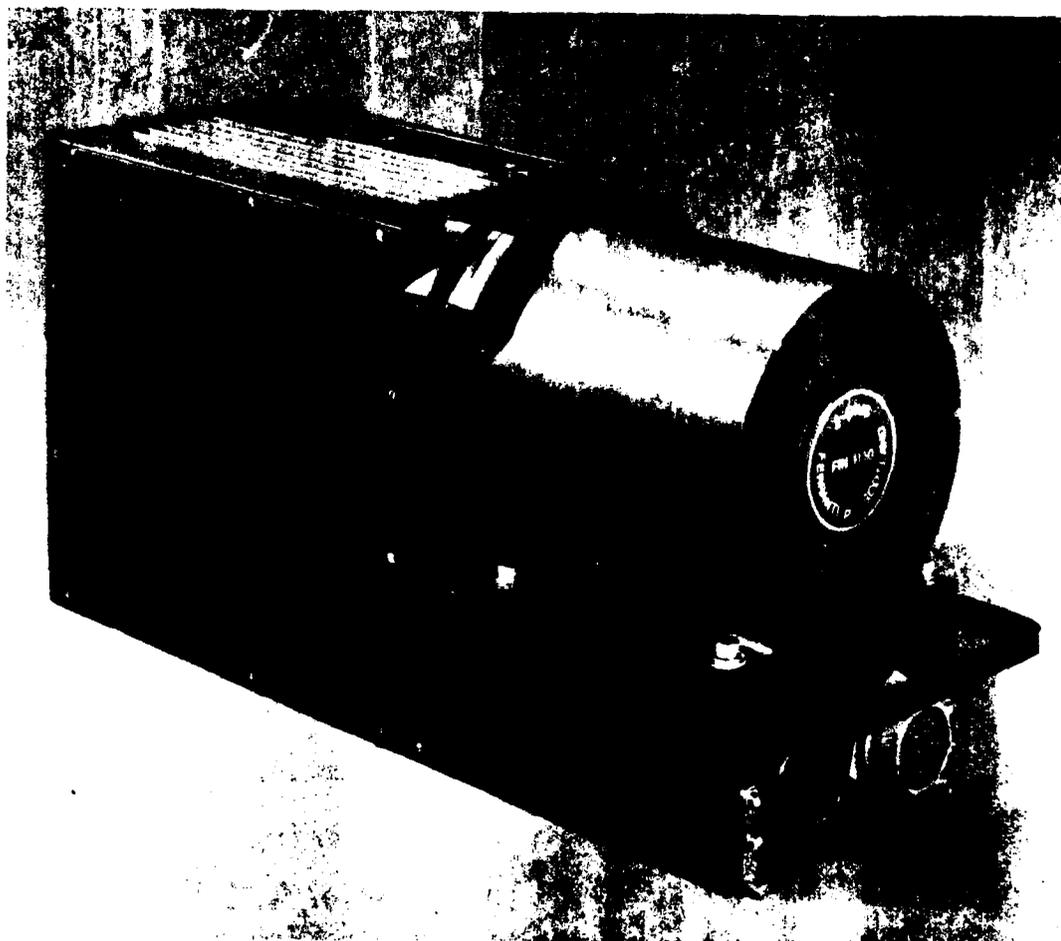


ILLUSTRATION 1

Ferranti FIN 1110 Two Gimbal Inertial Navigation System

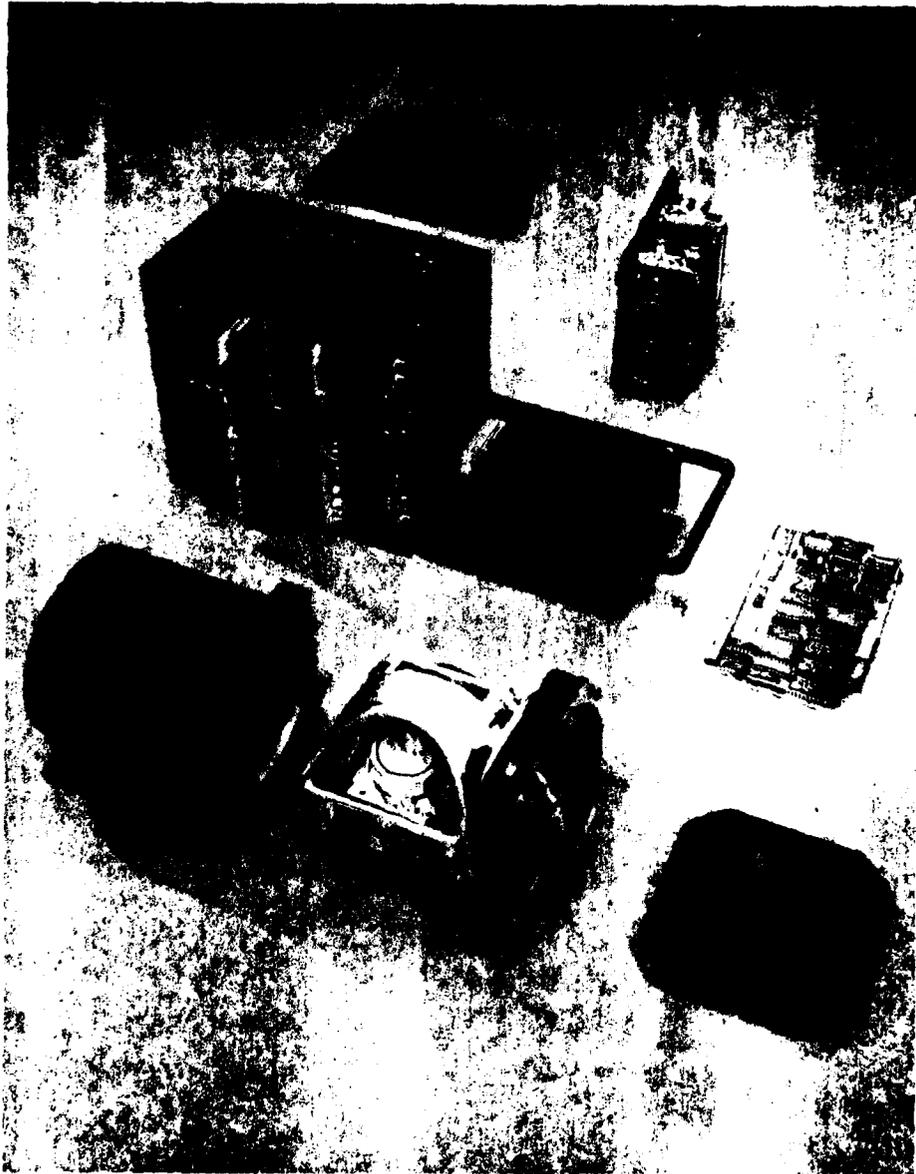


ILLUSTRATION 2
FIN 1110 Components

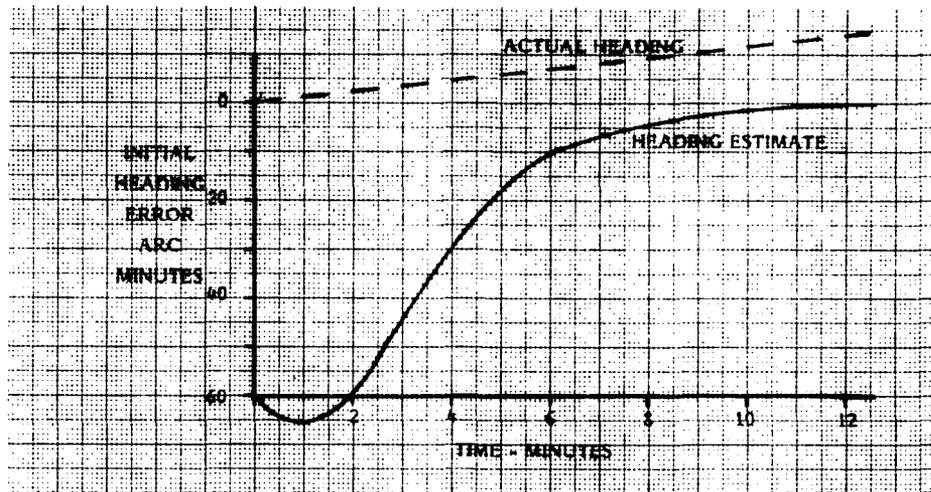


FIGURE 1 FIN 1110 - ALIGNMENT TIME - IN AIR ALIGNMENT

- Assumes - 1° Initial Heading Error
 - 1 Knot Doppler Error
 - $1.0^\circ/\text{hr}$ Azimuth Drift
 - 1 Knot Northerly Doppler Error

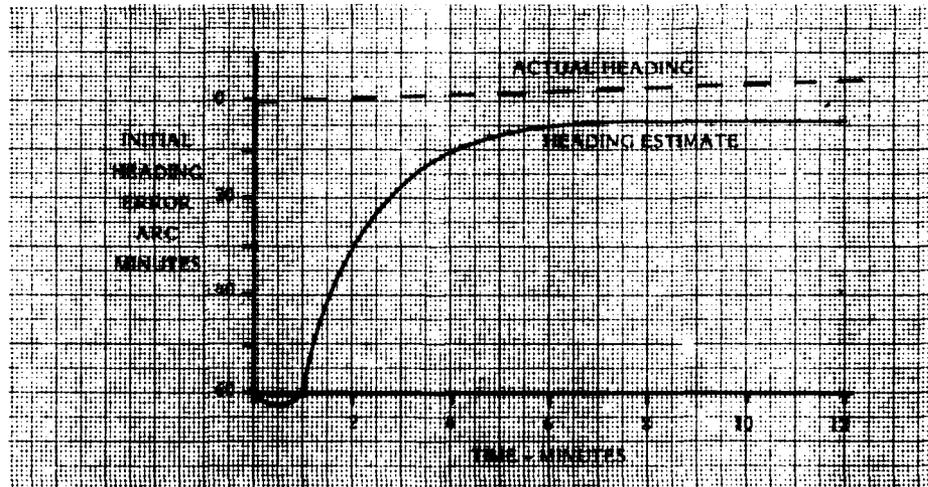


FIGURE 2 FIN 1110 - AZIMUTH ALIGNMENT TIME - GROUND ALIGNMENT

- Assumes - 1° Initial Heading Error
 - $0.2^\circ/\text{hr}$ Uncalibrated Azimuth Drift
 - 60° Latitude

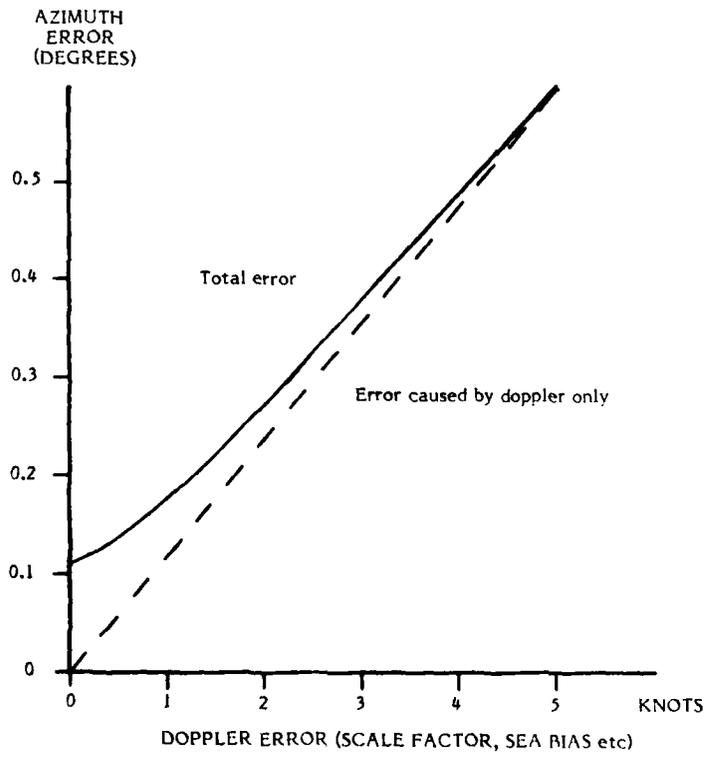


FIGURE 3
STEADY STATE AZIMUTH ERROR
Northerly Component of doppler error

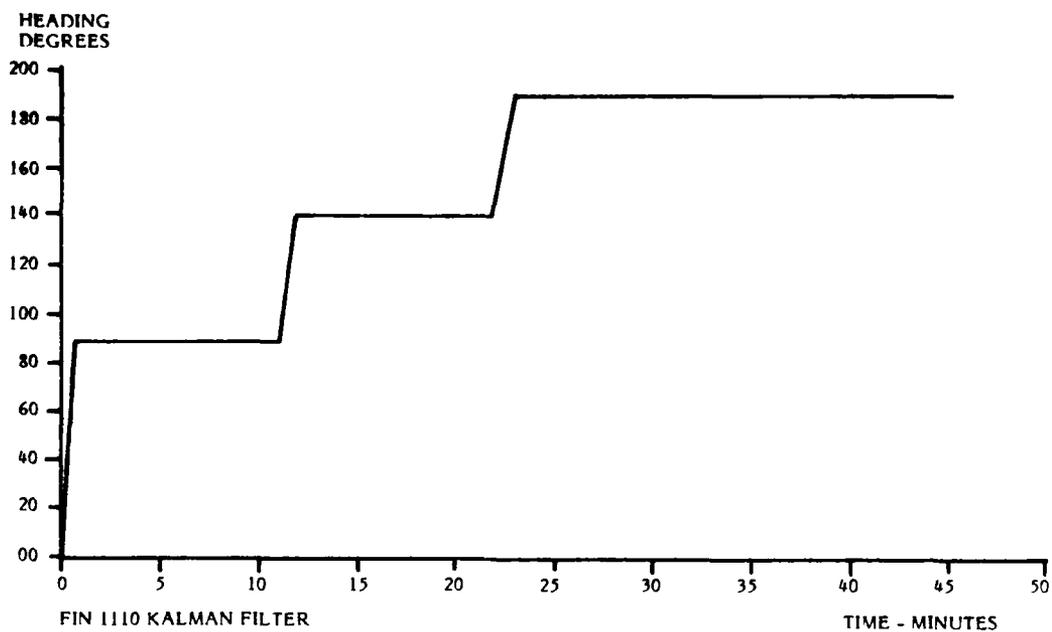


FIGURE 4-1

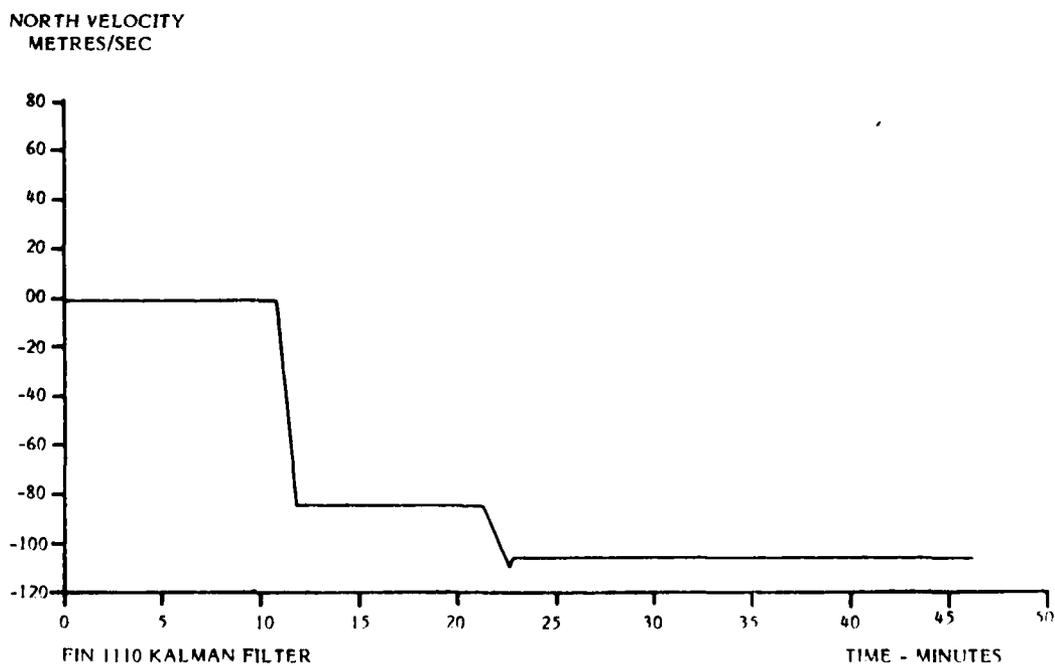


FIGURE 4-2

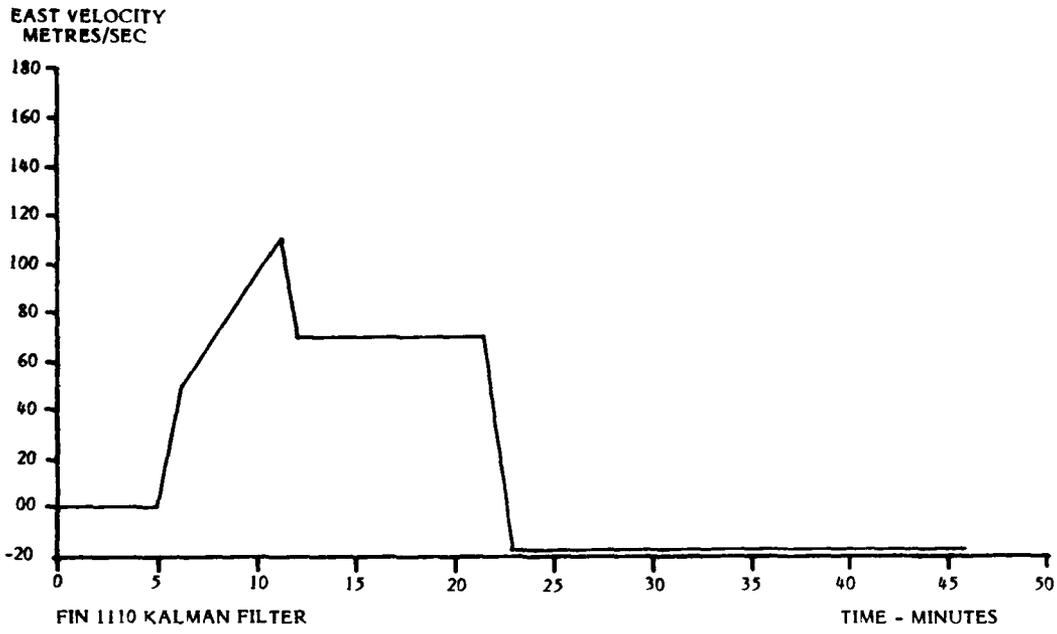


FIGURE 4-3

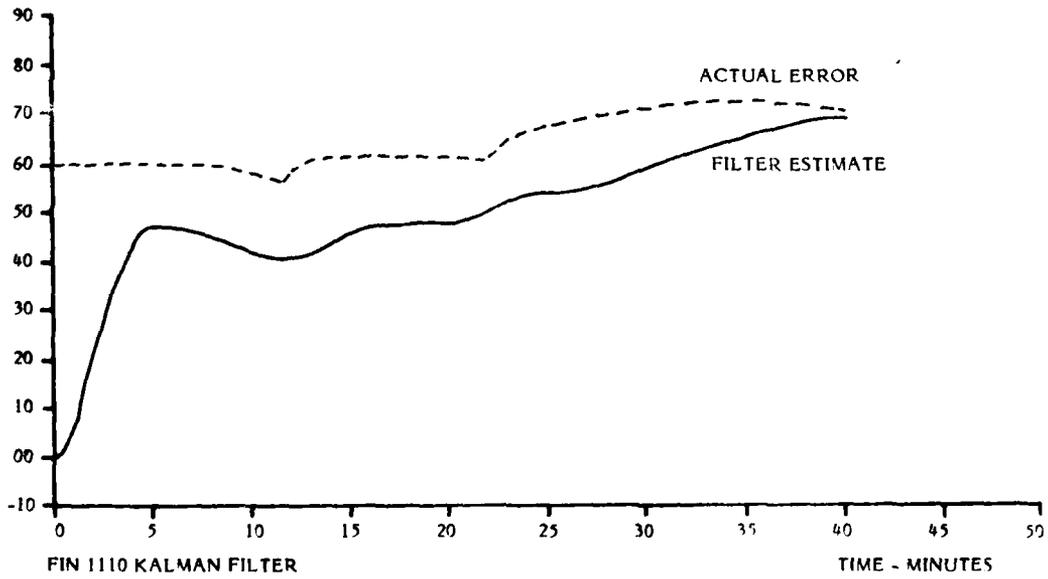


FIGURE 4-4

DOPPLER ALONG TRACK
SCALE FACTOR ERROR

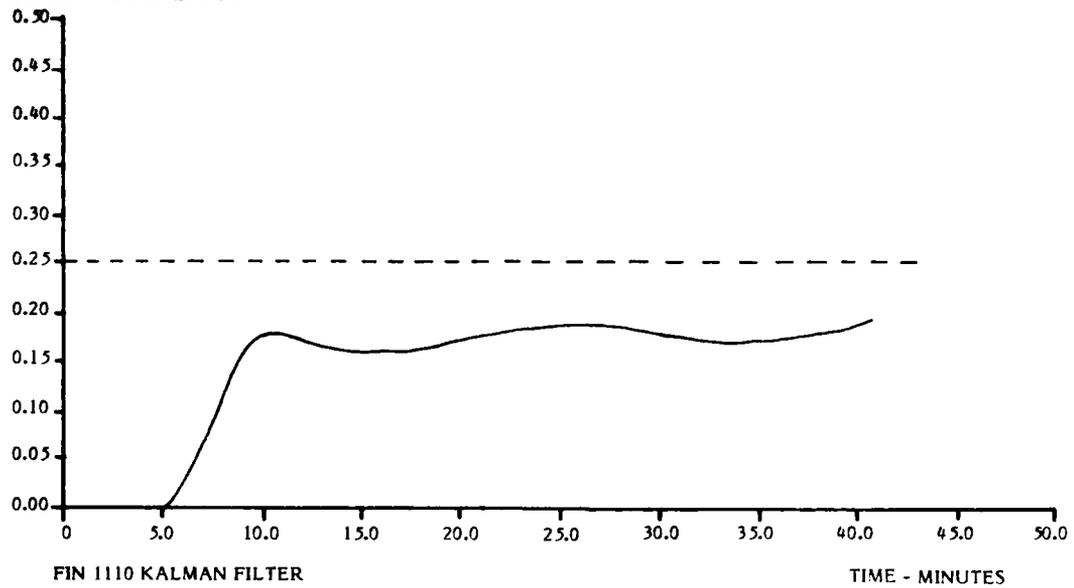


FIGURE 4-5

DOPPLER MISALIGNMENT ANGLE
ARC MINUTES

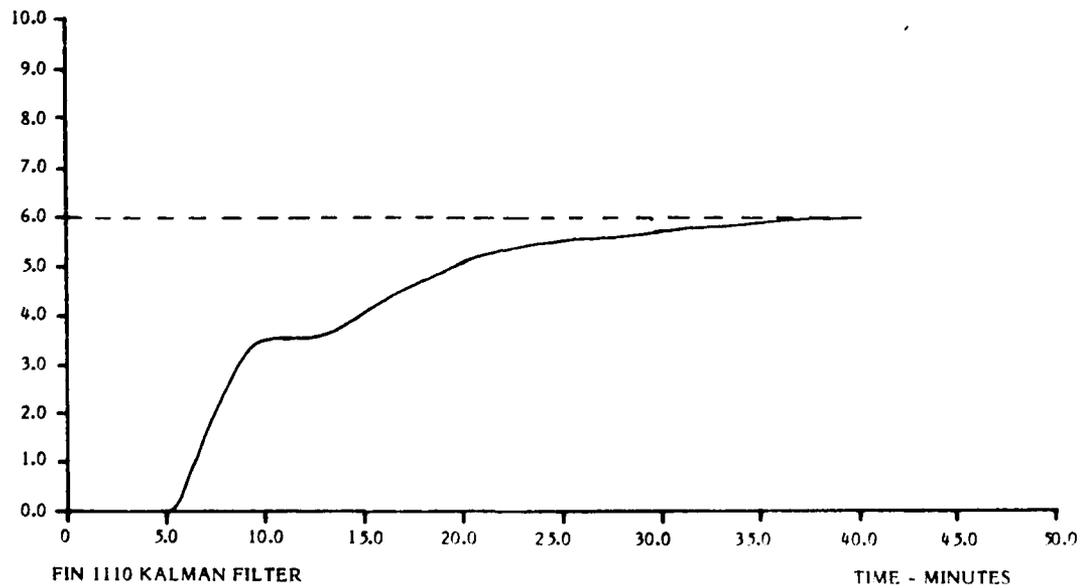


FIGURE 4-6

FIN 1110 FLIGHT TRIAL

NORTHING ERROR IN METRES

FLIGHT ONE

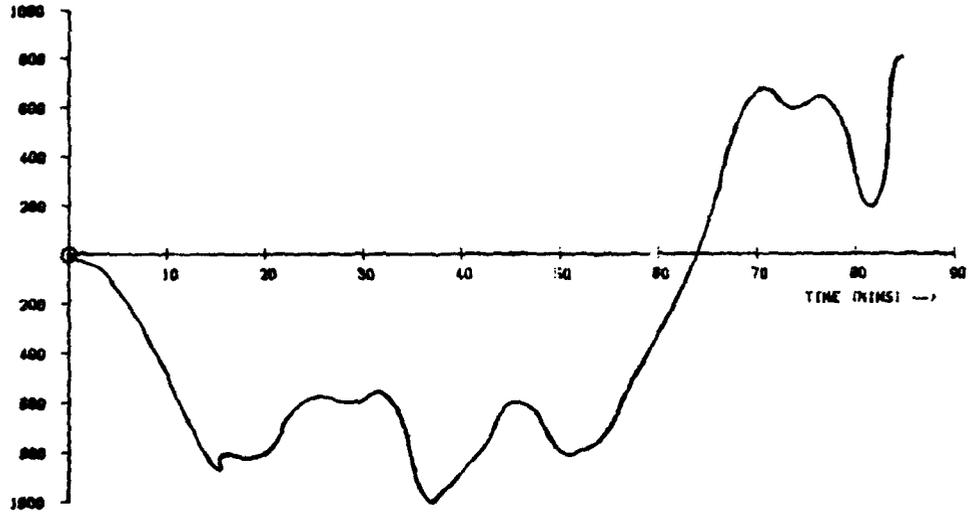


FIGURE 5A(1)

FIN 1110 FLIGHT TRIAL

EASTING ERROR IN METRES

FLIGHT ONE

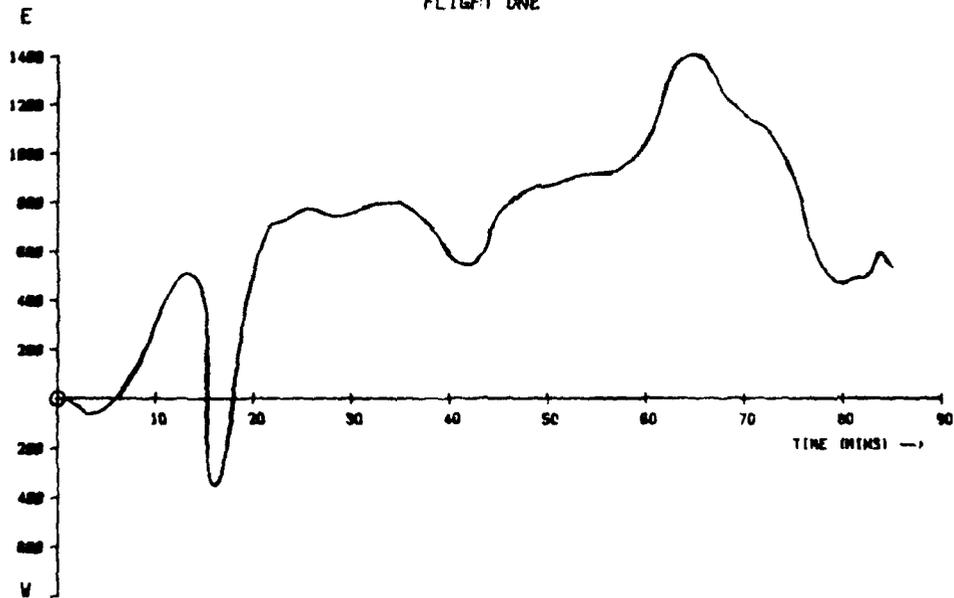


FIGURE 5A(2)

AD-A154 327

COST EFFECTIVE AND AFFORDABLE GUIDANCE AND CONTROL
SYSTEMS(U) ADVISORY GROUP FOR AEROSPACE RESEARCH AND
DEVELOPMENT NEUILLY-SUR-SEINE (FRANCE) FEB 85
AGARD-CP-360

4/4

UNCLASSIFIED

FEB 85

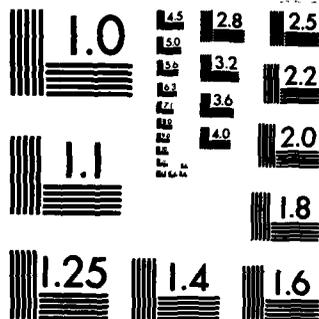
F/G 17/7

NL



END





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

FIN 111B FLIGHT TRIAL

NORTHING ERROR IN METRES

FLIGHT TWO

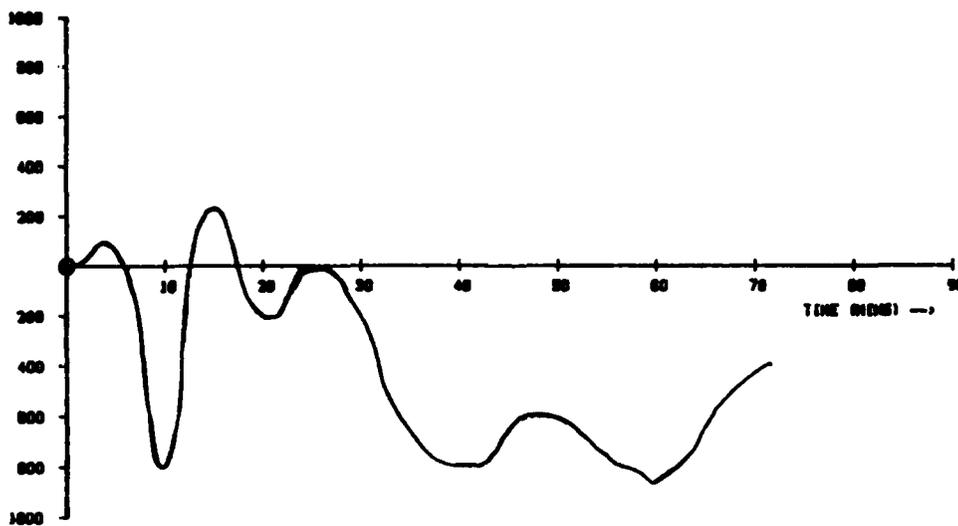


FIGURE 50(1)

FIN 111B FLIGHT TRIAL

EASTING ERROR IN METRES

FLIGHT TWO

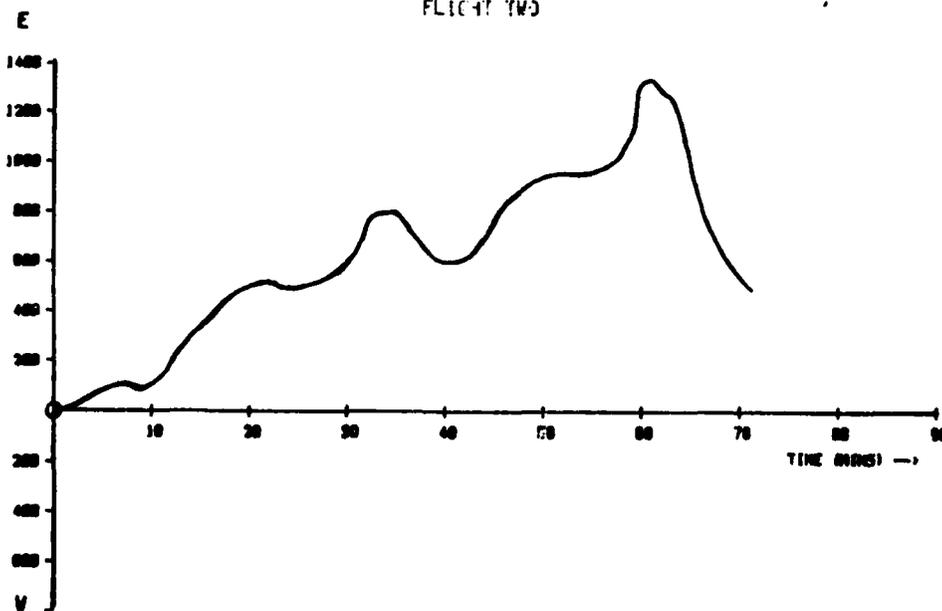


FIGURE 50(2)

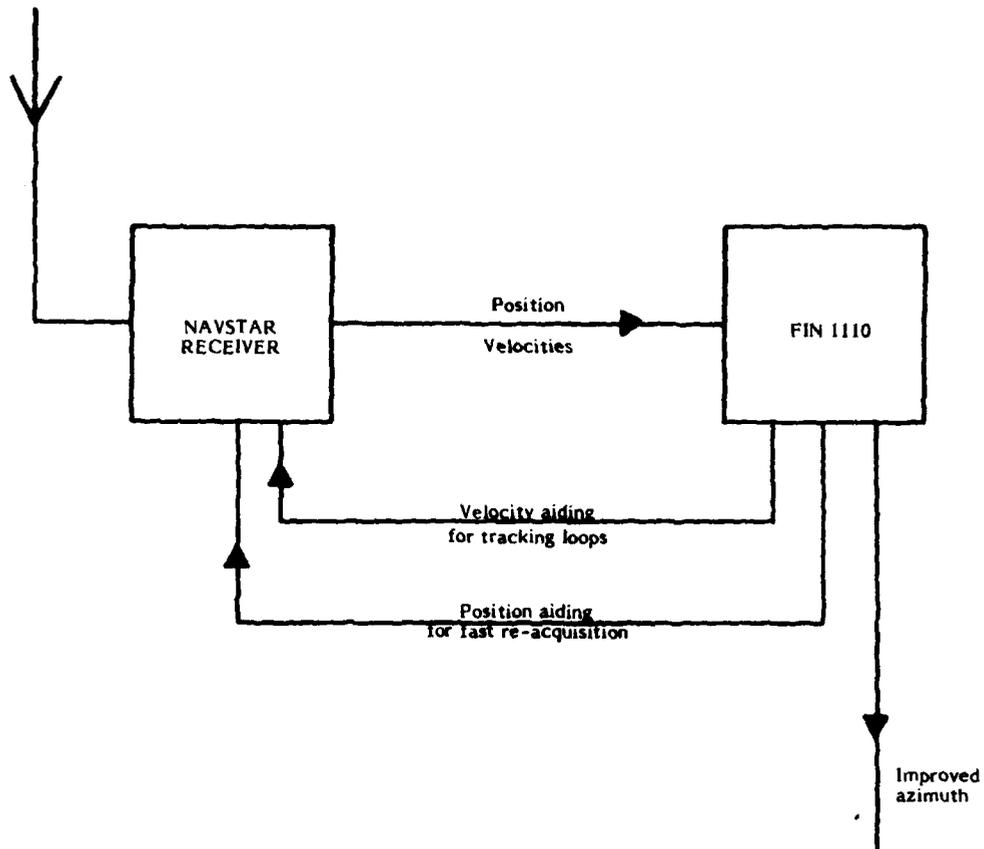


FIGURE 6
GPS/INERTIAL CONFIGURATION

COST-EFFECTIVE MISSILE GUIDANCE THROUGH DISTRIBUTED ARCHITECTURE
AND STANDARD PROCESSORS

by

H Gloerud

a.s Kongsberg Vaapenfabrikk
P.O. Box 25, N3601 Kongsberg
Norway

SUMMARY

The development of the Penguin MK3 missile was initiated by the Royal Norwegian Air Force as a modern stand-off weapon for the F-16 aircrafts. The guidance and control system of this missile is built upon the most modern principles using all digital electronics with distributed computational power. The system is separated in 6 main functions, each with its own powerful processor. High performance is achieved through the use of latest generation standard microprocessors and connected together by a high capacity data bus. Some of the results obtained from this development are described in this paper.

1 INTRODUCTION

The paper describes the guidance and control functions of the Penguin MK3 missile. This missile is an air-launched version of the Norwegian Penguin anti-ship missile. It is being developed for the Royal Norwegian Air Force for use with their F-16 aircrafts.

Figure 1 illustrates the exterior and interior of the missile. The target seeker is based on infrared detection. The control unit is all digital and contains the auto-pilot, trajectory generation and other functions. The inertial navigation platform is a semi strap-down platform (stabilized in roll axis) using two-axis dry tuned gyros. The platform provides 3-dimensional navigation and angular information. Missile flight control is provided by two pairs of canards actuated by a cold-gas powered hydraulic servo. The airframe is roll stabilized by electrically powered aileron actuators.

The operational requirements call for a missile to be launched from both high and low altitudes. It must be able to fly over terrain (mountains) and at a very low midcourse altitude. The target seeker concept also requires an airframe with high manoeuvrability.

Further the missile shall navigate along a specific trajectory and fly through (turn at) a predefined waypoint. Figure 2 presents an example of attack sequence using the radar mode. (Other modes are available).

To fulfill all these requirements it becomes necessary to have guidance and control functions with high performance.

2 DISTRIBUTED ARCHITECTURE

Early in the development program it was decided to distribute the digital guidance and control system into several independent functions with their own data processing capability. There were two main reasons for this:

- The data processing had to be split among several processors, since no known single micro-processor had the capability to do all processing needed.
- Development work could be done in parallel on independent functions. Specifications were established for each function and they were designed and tested separately, which contributed to shorter development time.

The complete guidance and control system is divided in five main functions plus a telemetry function for the test missiles. Figure 3 shows a block schematic of the functions and how they are distributed. The functions are:

- **Adapter.** This is located in a separate unit between the missile and the aircraft pylon. It interfaces the missile to the aircraft avionics multiplex bus and discrete signal lines. The function controls start-up and launch sequences. It also initiates an automatic test of the missile and keeps a test result record for maintenance purposes.

- **Control Function.** This function contains the missile autopilot and produces guidance commands to the canard and aileron actuators. These commands are control surface position commands. To give the missile good stability in flight, the actuators have a small signal bandwidth of approximately 30 Hz. Also to ensure adequate stability a high update rate of 200 Hz was chosen for the inner guidance loop computations. (Missile angular position and guidance commands). The outer guidance loop computes the correct trajectory, and deviations from it, on the basis of position data from the Navigation function and the altimeter.

Many of the autopilot algorithms are made adaptive with respect to flight altitude to account for variations in aerodynamic forces acting on the missile. Quick guidance response and good stability margins have been achieved.

The Control function also initiates various other actions during missile flight, e.g. motor ignition, altimeter transmission, seeker activation etc.

- **Navigation Function (INS).** The Penguin weapon operational concept requires a precise navigation function to ensure good target selection capability and a high hit probability even in confined coastal waters. The heart of this function is the inertial navigation platform (IMU). This is a 3-axis semi strap-down platform with stabilised roll axis. As output from the navigation processor the function provides body angles, position and velocity data.
- **Alignment Function.** To provide the required navigation accuracy this platform must be aligned with errors down in the milliradian range, velocity initialization must be better than 1 meter per second. Also excessive bias and scale factor errors of the gyros and accelerometers must be measured and corrected for. The only way to fulfill the above requirements on a fighter aircraft is to perform a transfer alignment with the aircraft INS as the reference. This means to compare the data from the aircraft INS and the data from the missile INS and then filter out the missile INS errors based on differences in data. The most convenient data to compare in this respect are velocity and angular data and this is done in the Alignment function. The Alignment function is implemented in a microprocessor located in the missile adapter as an 18 state Kalman filter using 3 velocity and 1 angular measurement (azimuth) as input. Figure 4 shows a block schematic of the alignment filter structure. The RELATIVE MOTION COMPENSATION block in the figure compensates for missile offset relative to the aircraft INS location when aircraft rotation occurs. Figure 5 shows a typical filter response presented as the standard deviation (milliradians) of the angular error estimates as a function of time (seconds). The pitch (PI) and roll (RO) deviations are quickly reduced from initial values of 8 mrad to approximately 1 mrad. However, the azimuth (AZ) deviation is significantly reduced when the aircraft performs a horizontal manoeuvre (3 g) which occurs after 60 seconds in the example shown.
- **Target seeker.** The seeker is based on the principle of infrared deflection. It contains advanced signal processing to give the missile high countermeasure resistance. Further several seeker modes are available. The seeker provides target position data to the autopilot.
- **Telemetry Function.** The test missiles are provided with a telemetry function. This function gather all data flowing between the functions, and transmit them for flight test purposes.

All these functions are tied together with the missile data bus. This is an 8 bit parallel (byte-wide) data bus with high data transfer capacity. Once normal data transfer mode is entered the bus uses no bus master. But data transfer is synchronized by a specific synch message from one of the bus terminals. The synch message is transmitted at a fixed frequency of 200 hertz. The various bus terminals may transmit messages on the bus in predetermined time slots relative to the synch message.

The bus is designed to have a maximum capacity of 500 kilo-byte per second, the capacity used in the missile today is 100 kilo-byte per second.

Data is transferred as messages of 4 to 50 byte length. 3 byte are used for identification and check.

The bus is tailored to the missile internal use and a maximum of 5 separate terminals may send messages on it. In addition other terminals may be listeners on the bus as is the Telemetry function. Depending on the nature of the subsystem (function) both dedicated terminal processor or circuits connected directly to the function's main processor is used as bus terminal.

3 STANDARD PROCESSORS

Each of the main functions mentioned above has its own powerful micro-processor. The tasks to be solved in each of these processors are such that only the most powerful 16/32 bit micro-processors available can do the job. The standard processor for these functions was selected in mid 1979. At that time very few alternatives existed.

The development program could not sponsor development of an own micro-processor. On the contrary it was important to select an established (or to become an) industry standard powerful micro-processor. This was important to keep costs down and be able to take advantage of the general technology development. The following is a list of other required characteristics reviewed before the final selection was made:

- Have second source.
- Available in military temperature range and quality.
- Have powerful support circuits.
- Also available in standard computer modules to be used in test and support equipment. This gives maximum standardization in software development.
- Support efficient use of high order language.
- Have powerful multi-user development support system.

Finally we found that the Motorola MC 68000 micro-processor satisfied our requirements and it was selected as the standard processor for the Penguin MK3 missile. It is used in all the functions described in chapter 2.

It was early decided to write as much as possible of the software in a high order language (HOL). The languages PASCAL and "C" were tested to find which was the most suitable for this purpose. From the compilers available at that time it was found that "C" gave by far the most efficient program code. The HOL "C" has therefore been chosen as the standard language. Although "C" is an efficient language it has been necessary to write a number of time-critical subroutines in assembly language, so a mix between "C" and assembly has been the result.

Standard computer modules with the same micro-processor have been very useful in test equipment. A dedicated bus monitor that can monitor and save data flowing on the missile data bus has been developed and is an example in this area.

4 RESULTS

The use of standard micro-processors and common solutions for various processor functions have helped to keep the time schedule in the development program. As a result the first live Penguin MK3 was fired almost exactly on the date planned more than 3 years earlier.

The results from this firing were very encouraging. The missile separated well, and navigated the trajectory as planned. Figure 6 and 7 are extracts from data collected from the first firing. Figure 6 indicates the guidance accuracy in altitude and figure 7 the missile pitch angle. Both figures indicate a missile guidance with good precision and stability.

The distributed architecture has made it easier to test the software of the individual functions. The total amount of missile software (approximately 200 kilobyte) have in this way been divided and structured and hence is easier to maintain. The various function's software could be tailored to the requirements of that particular function.

We were very early user of the MC 68000 micro-processor. That created some problems about availability of circuits, inadequate documentation and development support system malfunctions. Especially in the area of powerful support-circuits promises seldom came true. On the other hand the 68000 product family has really become the industry standard we expected.

5 CONCLUSION

It has been a successful approach to separate the system in selfcontained functions and have them developed in parallel.

Integration of these functions into a high performance guidance and control system has been done with few problems.

Growth potential is assured by the use of industry standard components from a family that is continually expanding with more powerful circuits.

The guidance and control system of the Penguin MK3 missile is probably not a truly low cost system. But high performance at moderate cost is achieved by the use of the distributed architecture and standard processors as described in this paper.

As a final remark and a trend example, it is worth mentioning that one Penguin MK3 missile with its adapter probably has more computing power than all computers in the F-16A together.

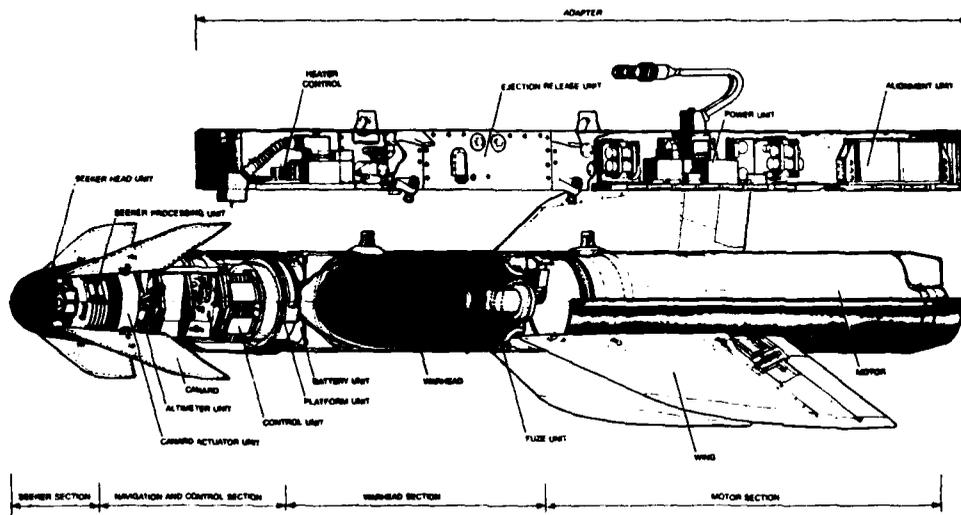


Figure 1 Outline drawing of the Penguin MK3 missile

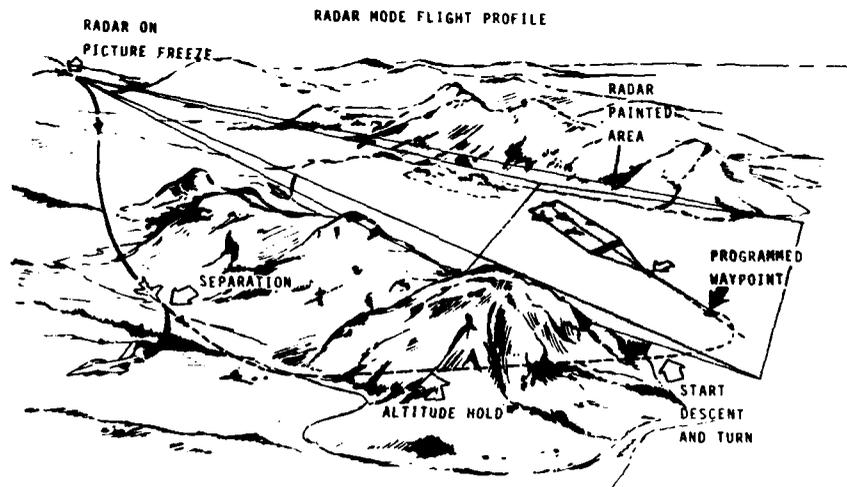


Figure 2 Illustration of a radar delivery mode sequence

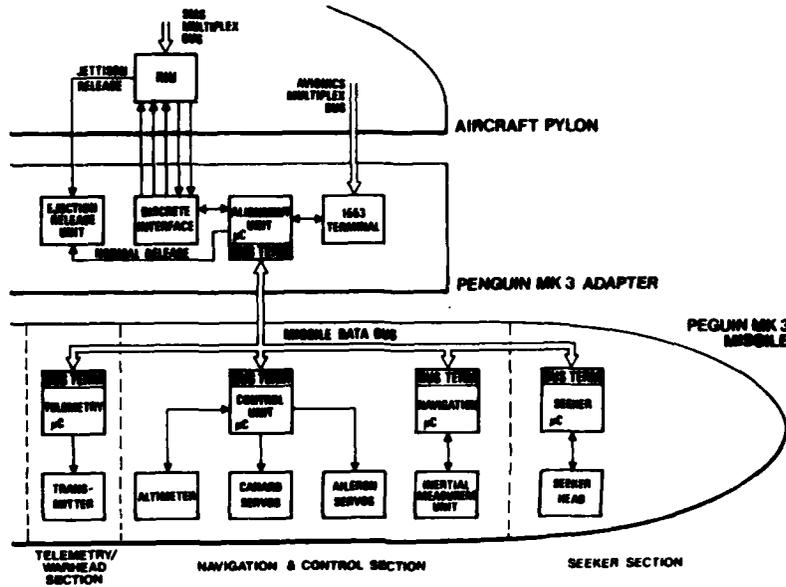


Figure 3 Block schematic of the distributed guidance and control functions

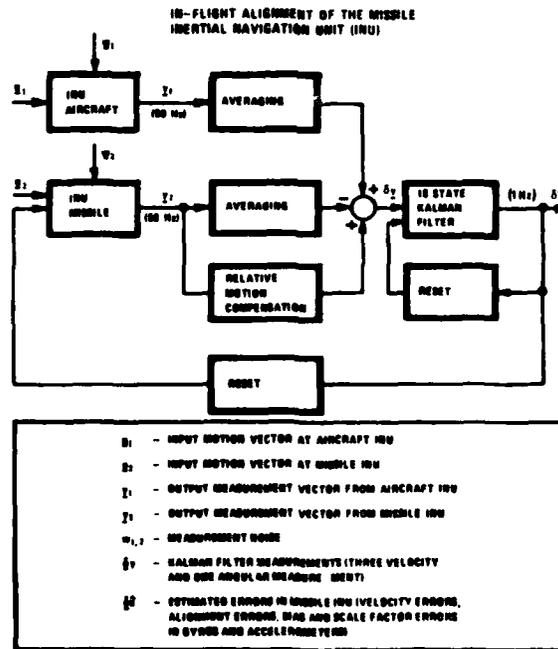


Figure 4 Block schematic of the INS alignment filter

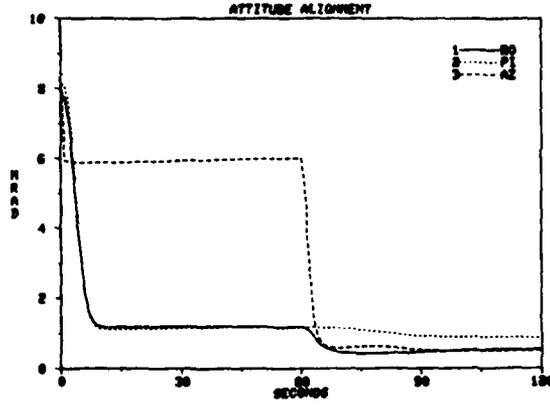


Figure 5 Example of alignment filter response presented as standard deviation of angular error estimates

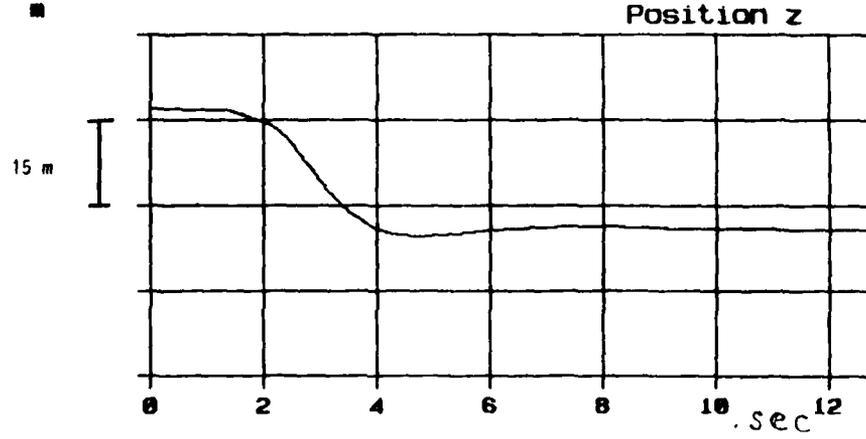


Figure 6 Missile altitude

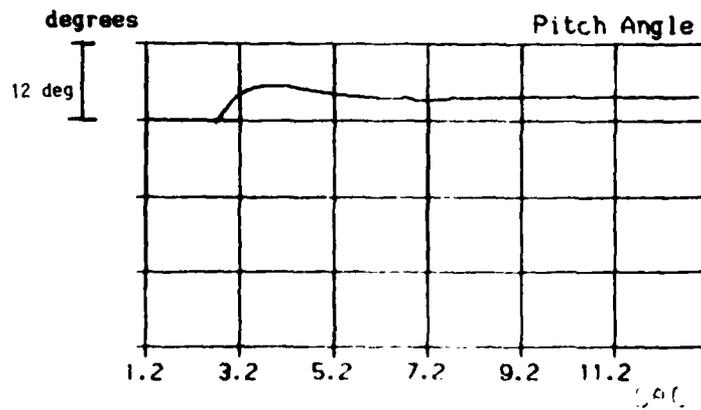


Figure 7 Missile pitch angle

| | | | |
|---|--|--|--|
| <p>AGARD Conference Proceedings No.360 Advisory Group for Aerospace Research and Development, NATO COST EFFECTIVE AND AFFORDABLE GUIDANCE AND CONTROL SYSTEMS Published February 1985 306 pages</p> <p>This volume contains the Technical Evaluation Report, the Keynote address and 24 out of the 26 papers presented at the Guidance and Control Panel 39th Symposium held at the Altin Yunus Holiday Resort Hotel, Çeşme, Izmir, Turkey from 16 to 19 October 1984.</p> <p>The papers covered the following topics: cost effectiveness models, systems configuration and design tools; low cost.</p> <p>P.T.O</p> | <p>AGARD-CP-360</p> <p>Cost-effectiveness Design to cost Cost requirements Reusable software Low cost platforms Wind modelling</p> | <p>AGARD Conference Proceedings No.360 Advisory Group for Aerospace Research and Development, NATO COST EFFECTIVE AND AFFORDABLE GUIDANCE AND CONTROL SYSTEMS Published February 1985 306 pages</p> <p>This volume contains the Technical Evaluation Report, the Keynote address and 24 out of the 26 papers presented at the Guidance and Control Panel 39th Symposium held at the Altin Yunus Holiday Resort Hotel, Çeşme, Izmir, Turkey from 16 to 19 October 1984.</p> <p>The papers covered the following topics: cost effectiveness models/systems configuration and design tools; low cost.</p> <p>P.T.O</p> | <p>AGARD-CP-360</p> <p>Cost-effectiveness Design to cost Cost requirements Reusable software Low cost platforms Wind modelling</p> |
| <p>AGARD Conference Proceedings No.360 Advisory Group for Aerospace Research and Development, NATO COST EFFECTIVE AND AFFORDABLE GUIDANCE AND CONTROL SYSTEMS Published February 1985 306 pages</p> <p>This volume contains the Technical Evaluation Report, the Keynote address and 24 out of the 26 papers presented at the Guidance and Control Panel 39th Symposium held at the Altin Yunus Holiday Resort Hotel, Çeşme, Izmir, Turkey from 16 to 19 October 1984.</p> <p>The papers covered the following topics: cost effectiveness models, systems configuration and design tools; low cost.</p> <p>P.T.O</p> | <p>AGARD-CP-360</p> <p>Cost-effectiveness Design to cost Cost requirements Reusable software Low cost platforms Wind modelling</p> | <p>AGARD Conference Proceedings No.360 Advisory Group for Aerospace Research and Development, NATO COST EFFECTIVE AND AFFORDABLE GUIDANCE AND CONTROL SYSTEMS Published February 1985 306 pages</p> <p>This volume contains the Technical Evaluation Report, the Keynote address and 24 out of the 26 papers presented at the Guidance and Control Panel 39th Symposium held at the Altin Yunus Holiday Resort Hotel, Çeşme, Izmir, Turkey from 16 to 19 October 1984.</p> <p>The papers covered the following topics: cost effectiveness models/systems configuration and design tools; low cost.</p> <p>P.T.O</p> | <p>AGARD-CP-360</p> <p>Cost-effectiveness Design to cost Cost requirements Reusable software Low cost platforms Wind modelling</p> |

high reliability guidance and control sensors; computational techniques and data processing; reduction of development and support costs; examples of cost-effective accomplishments and approaches.

ISBN 92-835-0373-2

high reliability guidance and control sensors; computational techniques and data processing; reduction of development and support costs; examples of cost-effective accomplishments and approaches.

ISBN 92-835-0373-2

high reliability guidance and control sensors; computational techniques and data processing; reduction of development and support costs; examples of cost-effective accomplishments and approaches.

ISBN 92-835-0373-2

high reliability guidance and control sensors; computational techniques and data processing; reduction of development and support costs; examples of cost-effective accomplishments and approaches.

ISBN 92-835-0373-2

B451
9

AGARD

NATO  OTAN

7 RUE ANCELLE - 92200 NEUILLY-SUR-SEINE

FRANCE

Telephone 745.08.10 - Telex 610176

**DISTRIBUTION OF UNCLASSIFIED
AGARD PUBLICATIONS**

AGARD does NOT hold stocks of AGARD publications at the above address for general distribution. Initial distribution of AGARD publications is made to AGARD Member Nations through the following National Distribution Centres. Further copies are sometimes available from these Centres, but if not may be purchased in Microfiche or Photocopy form from the Purchase Agencies listed below.

NATIONAL DISTRIBUTION CENTRES

BELGIUM

Coordonnateur AGARD - VSL
Etat-Major de la Force Aérienne
Quartier Reine Elisabeth
Rue d'Evere, 1140 Bruxelles

CANADA

Defence Scientific Information Services
Dept of National Defence
Ottawa, Ontario K1A 0K2

DENMARK

Danish Defence Research Board
Ved Idrætsparken 4
2100 Copenhagen Ø

FRANCE

O.N.E.R.A. (Direction)
29 Avenue de la Division Leclerc
92320 Châtillon

GERMANY

Fachinformationszentrum Energie,
Physik, Mathematik GmbH
Kernforschungszentrum
D-7514 Eggenstein-Leopoldshafen

GREECE

Hellenic Air Force General Staff
Research and Development Directorate
Hofargos, Athens

ICELAND

Director of Aviation
c/o Flugrad
Reykjavik

ITALY

Aeronautica Militare
Ufficio del Delegato Nazionale all'AGARD
3 Piazzale Adenauer
00144 Roma/EUR

LUXEMBOURG

See Belgium

NETHERLANDS

Netherlands Delegation to AGARD
National Aerospace Laboratory, NLR
P.O. Box 126
2600 AC Delft

NORWAY

Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
N-2007 Kjeller

PORTUGAL

Portuguese National Coordinator to AGARD
Gabinete de Estudos e Programas
CLAFE
Base de Alfragide
Alfragide
2700 Amadora

TURKEY

Department of Research and Development (ARGE)
Ministry of National Defence, Ankara

UNITED KINGDOM

Defence Research Information Centre
Station Square House
St Mary Cray
Orpington, Kent BR5 3RE

UNITED STATES

National Aeronautics and Space Administration (NASA)
Langley Field, Virginia 23365
Attn: Report Distribution and Storage Unit

THE UNITED STATES NATIONAL DISTRIBUTION CENTRE (NASA) DOES NOT HOLD STOCKS OF AGARD PUBLICATIONS, AND APPLICATIONS FOR COPIES SHOULD BE MADE DIRECT TO THE NATIONAL TECHNICAL INFORMATION SERVICE (NTIS) AT THE ADDRESS BELOW.

PURCHASE AGENCIES

Microfiche or Photocopy

National Technical
Information Service (NTIS)
5285 Port Royal Road
Springfield
Virginia 22161, USA

Microfiche

ESA Information Retrieval Service
European Space Agency
10, rue Mario Nikis
75015 Paris, France

Microfiche or Photocopy

British Library Lending
Division
Boston Spa, Wetherby
West Yorkshire LS23 3BQ
England

Requests for microfiche or photocopies of AGARD documents should include the AGARD serial number, title, author or editor, and publication date. Requests to NTIS should include the NASA accession report number. Full bibliographical references and abstracts of AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)
published by NASA Scientific and Technical
Information Branch
NASA Headquarters (NII-40)
Washington DC 20546, USA

Government Reports Announcements (GRA)
published by the National Technical
Information Service, Springfield
Virginia 22161, USA



Printed by Specialised Printing Services Limited
40 Chigwell Lane, Loughton, Essex IG10 3LZ

ISBN 02 835-0373 2

END

FILMED

7-85

DTIC